

nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

Sea change

*Seagrass genome reveals key adaptations
of a flowering plant that returned to the ocean*

PAGES 290 & 331

GRAVITATIONAL WAVES

ABOUT SPACE-TIME

Historic discovery opens
new window on the Universe

PAGE 261

EVOLUTION

THE FIRST PREDATORS

The role of ecology in the
Cambrian radiation

PAGE 268

EMISSIONS

GREENER SHIPPING

Reduce the environmental
impact of sea freight

PAGE 275

NATURE.COM/NATURE

18 February 2016 £10

Vol. 530, No. 7590

07



THIS WEEK

EDITORIALS

MICE Tasting menu points to environmental differences **p.254**

WORLD VIEW Push innovation science onto the agenda **p.255**

ROACH ROBOT Squashy survivors inspire flattening device **p.257**



Green-sky thinking

Environmental agencies must go much further in regulating aircraft emissions if they want to make a real difference.

Attitudes towards flying say a lot about someone's view on global warming. A hardy bunch of committed worriers take the train instead, whereas others still celebrate the jet-set lifestyle as a sign of success. Then there are those who fly, but feel guilty about doing so.

Aviation has become a symbol of the world's reluctance to make serious efforts to tackle climate change — perhaps unfairly, given its relatively slight (although growing) contribution to the global-warming problem. On an individual level, those who travel by air leave gigantic carbon footprints, governments continue to invest in runways and airports, and the industry remains focused on growth.

Most international frameworks to tackle carbon emissions struggle to include aviation. When the European Union tried to encompass emissions from international aviation in its emissions-trading scheme in 2012, it met with widespread protest from the industry and governments. Instead, the International Civil Aviation Organization (ICAO) — the United Nations body that oversees the skies — agreed to come up with its own measures.

The world saw the initial results of the ICAO's work last week, when the organization proposed a new global carbon dioxide standard for aircraft (see page 266). It was hardly an inspiring achievement. The proposed regulation, which is expected to be adopted later this year, is complex, but the gist is that all new aircraft would need to meet minimum fuel-efficiency standards by 2028. The ICAO says that the rule will guarantee reductions in CO₂ emissions. This may be true, but it is misleading.

An independent assessment by the International Council on Clean Transportation (ICCT) suggests that new aircraft would emit on average 4% less CO₂ when the measure takes full effect. However, each generation of new aircraft is already made to be more fuel efficient than the last, and the same independent assessment highlights that aircraft manufacturers are likely to achieve an efficiency improvement of more than 10% by the time the new standard kicks in, effectively rendering the rule redundant.

Still, the most notable thing about the global standard will be that it exists. It is both a precedent and a tool that could one day be used to push the industry further than it would go of its own accord.

Individual countries could yet adopt stricter regulations. Last year, the US Environmental Protection Agency (EPA) issued an 'endangerment' finding for aviation emissions, which represents the first step in a regulatory process under the country's Clean Air Act. The EPA is expected to finalize its finding in the coming months, and then it could launch its own regulatory proposal. The agency could, and should, go well beyond the ICAO standard on new aircraft, and introduce rules for existing aeroplanes.

The EPA will not be able to complete this process before President Barack Obama leaves office, so it will be up to whoever is elected president in November to follow through. Given the general opposition by US conservatives to any kind of action on climate change, there is little

hope of getting a strong regulation from a Republican administration. Moreover, whatever the EPA proposes will surely be challenged in the courts, which can be fickle and unpredictable, as evidenced by the Supreme Court's decision last week to block implementation of Obama's power-plant regulations pending the outcome of a legal challenge. But one thing is clear: the EPA must act on flights, and environmentalists will surely take the agency to court if it does not.

"Now is a good time to invest in a much cleaner future."

Nor is the ICAO's work done. The body will now address a plan to halt emissions from international aviation at 2020 levels. This is crucial because international aviations already account for roughly 1.4% of global CO₂ emissions and are currently unregulated. Even the global climate agreement signed in December in Paris neglected to account for emissions from aviation or from international shipping, which is responsible for nearly 1.8% of the world's CO₂ emissions (see page 275).

Zero-emissions aircraft are not likely to be flying any time soon, so the key to the ICAO's idea is the use of carbon offsets. It is probable that some kind of fee would be levied on international flights to pay for emissions reductions elsewhere. But there is scope to go further on cleaner aircraft too.

Airlines are currently reaping profits thanks to the collapse of the oil market, which has lowered fuel prices across the board. Despite opposition from the aviation industry to strong emissions rules, now is a good time for it to invest in a much cleaner future. ■

Back to Earth

Success against cancer need not deliver the Moon.

When John F. Kennedy pledged in a 1961 presidential speech to land a man on the Moon and return him safely to Earth, he launched more than a space programme. He introduced the ultimate metaphor. Today, moonshots no longer need to shoot for the Moon. They can signify merely the launch of a grand effort fuelled by bold ambition that will elevate society to some new heights.

The latest is the US Cancer Moonshot, a US\$1-billion plan, to be spearheaded by vice-president Joe Biden, that aims "to eliminate cancer as we know it".

The project and the promised investment are welcome indeed. The name and the rhetoric less so, and not just because they are so unoriginal — moonshots and Apollo programmes have been launched

in recent years on everything from renewable energy and neuroscience to an assortment of Google X pet projects and at least three efforts to fight cancer.

Perhaps the United States was due for another national promise to cure cancer: the last — a 2005 pledge by Andrew von Eschenbach, then head of the US National Cancer Institute — was scheduled to have vanquished the disease by 2015. This followed then President Richard Nixon's 1971 pledge to use \$100 million to cure cancer. To be sure, pledges to cure cancer have a long history of succeeding in one respect: fundraising. But the idea that \$1 billion could eliminate cancer is misleading, and only becomes more so as each passing year reveals more about the full challenge of fighting the disease. With apologies to Biden, the more we 'know it' the harder it becomes to think we can 'eliminate cancer'.

Today, we have a clearer view of cancer's complexity. The sequencing of tumour genomes has revealed heterogeneity not only among cancers and patients, but in a single tumour. Within those complex mixtures of cells can lurk mutations that give rise to drug resistance. Therapies against cancer-causing mutations have been transformative for some patients in the short term, but nearly always fail in the long term as resistant cells reseed the tumour.

Real progress is being made, little by little. Chemotherapy cures more than 85% of children with acute lymphocytic leukaemia, for example. And for a small number of patients with various cancers, new immunotherapies have produced remissions so prolonged that doctors have begun to whisper the word 'cure'. But although combinations of these therapies could hold the key to expanding their success, testing combinations in clinical trials is complex — and quite likely to cost more than \$1 billion. As with everything, the more successful we get, the harder it is to improve. For many cancers, a more reasonable aim might be to turn them into chronic, manageable diseases.

In statements and conversations, Biden has acknowledged this complexity, and has even reportedly expressed regret for choosing the moonshot theme. It is unfortunate that sound bites from Biden and

the White House continue to back its simplistic framework. Repeated invocations of these bold but doomed quests to cure cancer in a decade, or with a given sum of money, feed public cynicism about the value and potential of science. And setting an unreachable goal plays down the tremendous progress that cancer researchers have made.

Details of the latest cancer moonshot are sketchy; Biden is still gathering input from the country's scientific glitterati. From what we know, he

“Setting an unreachable goal plays down the tremendous progress that cancer researchers have made.”

hopes to double the pace of cancer research by breaking down the barriers — logistical and cultural — that keep researchers from sharing data. This can include having electronic medical records that talk to one another, encouraging collaboration, and developing central repositories that can handle big data. Some of these issues are already being tackled: the National Cancer Institute, for example, is putting together a large database that aims to

unite disparate data sets, along with clinical information and the privacy concerns entailed, in one place.

There is also no guarantee that Obama's \$1-billion request will come to pass. Congressional leaders have pledged to ignore the president's budget request. And Obama sought to establish the funding for the cancer moonshot using an unusual approach that would circumvent the usual congressional funding process. Congress is unlikely to sign up to that. But there is hope that the programme will survive in some form: this Congress has a soft spot for medical research, and Biden is an authority on the art of congressional compromise.

Let us hope it will. Biden's early vision of the programme, if executed well, has the potential to be high-impact. Cancer research is in the middle of a revolution, and may be on the brink of even greater success. The US Cancer Moonshot has the potential to build on this momentum. The project does not need to mislead the public, and damage its trust in science, in the process. ■

Chow down

Scientists should pay more heed to the varying effects of diet and environment on animal work.

Japanese scientists last year reported the results of an extraordinary experiment in animal nutrition. The mice they worked with could well have been the best fed in the history of research — not in terms of quantity of food, but in its quality.

On a typical day, one group of mice got to eat mixed rice with dried whitebait and green seaweed flakes for breakfast, together with cooked beans and miso soup containing the root vegetable taro and Japanese mustard spinach. Another group got bacon and eggs, toast and fluffy boiled potatoes. Lunch for one group could be simmered pumpkin and ground chicken, with a portion of cucumber and wakame seaweed with vinegar dressing. A different group of mice got a hamburger and salad.

Dinner was selected by the scientists from dishes including prawns with chilli sauce, Sichuan-style bean curd, fried Japanese horse mackerel, white radish and shimeji mushroom soup, sake-steamed clams and steamed Japanese seerfish. The mice ate from that kind of menu every week of their lives. There was no pudding, but probably no complaints.

The reason for all this gourmet cuisine was to recreate the typical Japanese diet from decades past, and to examine its impact on health. The long life expectancy of people in Japan has been attributed to the benefits of *Nihon shoku* or traditional diet, involving fermented foods that seem to boost the protective effect of harmless microbes on and in the body. As the food available in Japan has become increasingly Westernized, the effects on health are being questioned.

Hence the mouse study. Each group was given dishes from recreated daily menus for a typical Japanese household in 1960, 1975, 1990 and 2005. The food was ground up and fed to the animals along with their regular meals of 'standard laboratory chow'. As the scientists suspected, the animals that were fed the older, more traditional diets lived and prospered for longest (K. Yamamoto *et al.* *Nutrition* **32**, 122–128; 2016).

There are two things to note. The first is the large contribution that the environment — in this case diet — can make to health. The second is that such experiments enable the impact of environment on health to be assessed in ways that are simply not possible for human trials.

Light, heat, food, company, exercise, distractions, stress — all are at the fingertips of scientists who set up mouse experiments. Subtle changes in any of these can lead to profound, and potentially useful, discoveries about how health is changed by external factors. Research has probed, for example, how giving mice tunnels, stairs and wheels to play with can alter how female mice interact with their young, which in turn alters the brain development of the offspring (T. Begensic *et al.* *Neurobiol. Dis.* **82**, 409–419; 2015). There is some evidence that modern, sterile, individually ventilated cages — used to minimize the effects of environmental factors such as disease — are quieter and less smelly for the mice, which reduces stimulation of those sensory systems.

Given that we know that environment affects the outcome of experiments, it is surprising that we don't know more about the environmental set-up of other studies — those that test the impact of a potential medical treatment, for example. As we report in a News story on page 264, many researchers who use mice do not even know

the content of standard lab chow and how it may change from study to study. As scientists hunt for sources of irreproducibility in their research, variation in living conditions — and how to reduce it — deserves more attention. ■

© NATURE.COM

To comment online,
click on Editorials at:
go.nature.com/xhunqy



Put innovation science at the heart of discovery

The success rate of discoveries would be improved if we could find out how to innovate, argues Andrew Kusiak.

Innovation is being talked about everywhere. The US Senate is working on a biomedical innovation bill. Australia's main funding agency has just announced that it will cut hundreds of climate scientists as part of its National Science and Innovation Agenda. The National Council of Science Museums in India will add Innovation Hubs at its centres. More and more organizations are using innovation in their names and brands. Innovation is a central plank of national and local policies and it consumes billions of dollars of investment worldwide. Yet the evidence base for these innovation efforts is close to nothing. We simply do not know how the innovation happens. We should do more to find out.

Innovation is commonly confused with invention and creativity. Creativity is the ability to generate original ideas, concepts and objects. It spurs invention, which is most evident in the areas of technology and business. Artists enjoy creativity, whereas engineers and scientists focus on inventions. But innovation demands a third ingredient: market success.

History contains examples of the rapid transformation of creativity into market success: Picasso managed to earn an income from his creations, and Disney's theme parks are lucrative tourist attractions. But disruptive innovations — those that have transformative impacts, such as the steam engine or the Apple iPhone — are rare.

The path to innovation is currently more art than science. That might explain why it is shockingly inefficient: the chance of an invention attaining enough commercial or social success to be recognized as an innovation reaches no more than low single percentages. In the United States' Small Business Innovation Research programme, a very low proportion of grants results in a viable economic activity, product or service. In markets that are saturated, such as those of mobile phones or medical discoveries, the success rate is even lower.

Governments want innovation that not only transforms the industry, but also offers solutions to the 'grand challenge' problems that the world faces: alternative energy sources, mitigating climate change, eliminating poverty and improving health care and security. Many research programmes in these and other areas claim to be innovative, because they seek and apply new approaches to a specific problem. Others seem to believe that the research itself is innovative because it produces new findings, or that the results will inevitably lead to innovative outcomes.

But it is not that simple. There is no deep understanding of the innovation process, which is complex and has not been well captured or formalized. There is no unified theory or reliable model for innovation. There is no innovation science.

How could the science of innovation happen? There are several areas in which research could initiate and potentially formalize it: for example, the study of patents or creative individuals such

as musicians and painters. We could use results from these studies to conceptualize and model the innovation process from generalizations identified across the engineering, arts, science and social domains.

The task is complex and enormous. The first step could be to look for shared factors and to devise rules and hints that support innovation. It might help to look backwards. Firms and individuals often claim that they have learned from mistakes, but how many analyse failure systematically? Patent libraries are packed with submissions that never get used, and many research programmes and clinical trials do not lead to success. Analysis of these failures could help others to succeed, and could contribute to an understanding of what drives innovation.

Another backward-looking approach is imagining the best, then scaling it back to reality. Imagine an item of office furniture with all-encompassing functionality: it fulfils all the needs of the worker, but

also changes colour according to the weather and adapts to the height and weight of the person. Limitations of technology, prohibitive cost and anticipated market response are then used to scale it back into marketable items: a chair whose height is adjustable and a desk that can be set at two height levels, and both would be available in different colours. Computer printers were innovated in a similar way by incorporating functions beyond printing.

A building block of innovation science is connecting seemingly unrelated ideas. We are flooded with discoveries in isolated domains. Making quick connections between, for instance, biology and technology, could lead to bigger ideas and redirect research and development.

Innovation-science researchers must develop models of the market and products to predict successful outcomes. These models could be based on emerging evolutionary computation, and would be developed, validated and tested using streams of data, such as consumer interests and preferences.

Over the long term, private foundations should establish a global initiative at a scale similar to that of the Breakthrough Energy Coalition spearheaded by Microsoft co-founder Bill Gates. (In fact, that coalition could itself greatly benefit from innovation science.)

In commercial terms — comparing investment to output — the innovation process might have a failure rate of 99%. That would simply not be tolerated by any other commercial enterprise. Mainstream industry has moved to six-sigma programmes and beyond, barely tolerating one or two errors in a million. Governments and scientists should focus less on discussing various forms of innovation and more on how to innovate. ■

THERE IS NO DEEP UNDERSTANDING OF THE INNOVATION PROCESS.

◀ **NATURE.COM**
Discuss this article
online at:
go.nature.com/mewfri

Andrew Kusiak is professor of mechanical and industrial engineering at the University of Iowa in Iowa City.
e-mail: andrew-kusiak@uiowa.edu

RESEARCH HIGHLIGHTS

Selections from the scientific literature

MICROBIOLOGY

Bacterial version of an eyeball

A freshwater bacterium can sense the direction of light by acting as a microscopic lens.

Spherical cyanobacteria called *Synechocystis*, which harvest energy from light, use protein appendages to pull themselves over wet surfaces towards light sources. A team led by Conrad Mullineaux at Queen Mary University of London found that the cells act as tiny lenses that bend and focus light. When the team illuminated one side of the cell, a bright spot appeared at the opposite end. Simulating this spot with a laser beam caused the *Synechocystis* cells to move away from the spot, towards the perceived source of light.

Light-sensing proteins embedded in the cellular membrane trigger the bacteria to move towards light, the researchers suggest.

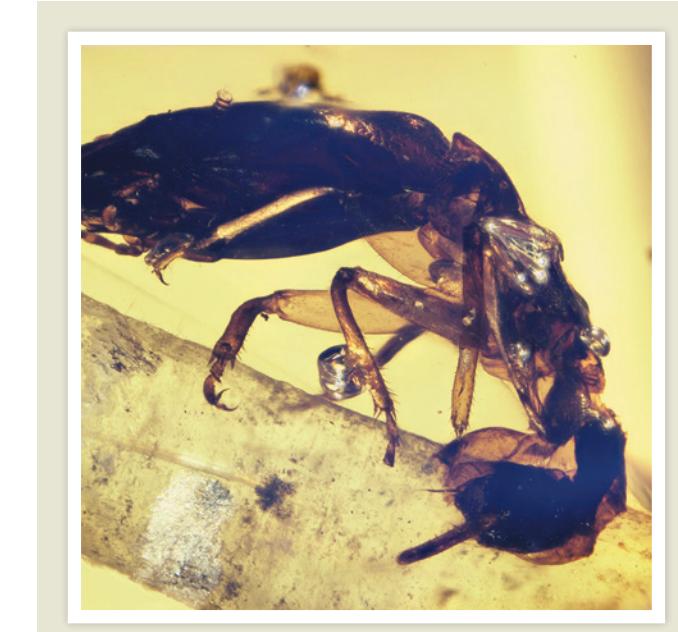
eLife <http://doi.org/bcgd> (2016)

NEUROSCIENCE

Brain circuit for loneliness

A neural circuit at the base of mouse brains drives a loneliness-like state and motivates the animals to seek company.

Kay Tye at the Massachusetts Institute of Technology in Cambridge, Mark Ungless at the Medical Research Council's Clinical Sciences Centre in London and their colleagues found that connections between neurons in the circuit were stronger in mice that were separated from their cage mates than in those that were grouped together. Those neurons then fired more frequently when isolated mice were put in a cage with an unfamiliar mouse, compared



PALAEONTOLOGY

Termites had first castes

The separate castes of social insects — queens, workers and soldiers — first appeared in termites at least 100 million years ago, according to newly discovered fossils.

Although evolutionary studies have suggested that termites were first to evolve castes, there was very little fossil evidence for this. Now, Michael Engel at the University of Kansas in Lawrence, David Grimaldi at the American Museum of Natural History in New York and their colleagues report the discovery of six termite species fossilized in amber from Myanmar, which show evidence of distinct castes.

The 100-million-year-old fossils include two new species, which the authors call *Krishnatermes yoddha*, with queens, workers and soldiers (pictured), and *Gigantotermes rex*, whose 2-centimetre-long soldiers are among the largest ever reported.

Curr. Biol. <http://doi.org/bch6> (2016)

with animals that had not been isolated. When the scientists inhibited the neurons with light, the isolated mice showed less interest in the stranger. Activating those neurons caused the animals to actively seek other mice.

The circuit was more responsive in socially dominant animals.

Cell 164, 617–631 (2016)

the risk that cancer will come back. To address this, Dmitri Lapotko, now at the medical-technology firm Masimo in Irvine, California, and his colleagues injected gold nanoparticles into tumour-bearing mice before surgery. The particles, which have cancer-specific antibodies on their surfaces, were taken up by cancer cells. After removing the main tumour, the researchers heated up the nanoparticles using a short laser pulse, causing nanobubbles to form and explode only in the cancer cells, destroying them without harming normal tissue. The explosions generated a detectable acoustic signal.

The team spotted small numbers of these cells during surgery that would otherwise have gone unnoticed. After the surgery, no tumours regrew in any animals in which residual cancer was removed, whereas more than 80% of the mice that had standard surgery died of recurring tumours.

Nature Nanotech. <http://dx.doi.org/10.1038/nnano.2015.343> (2016)

MATERIALS

Power from water and graphene

Chemists have generated electricity from water by passing it through a material containing atom-thick sheets of carbon.

Liangti Qu at the Beijing Institute of Technology and his colleagues developed a 3D structure made from graphene oxide that had holes big enough to let moisture pass through them freely. The water molecules reacted with oxygen-containing groups in the graphene oxide, dissociating to form hydrogen ions. The oxygen groups were distributed unevenly in the material, with more at the bottom than the

BIOENGINEERING

Exploding bubbles kill cancer cells

A technology using tiny exploding bubbles inside tumours could one day help to mop up cancer cells during surgery.

Surgeons lack the tools to detect microtumours during cancer surgery, increasing

top, resulting in a large-enough flow of ions to generate electric power.

The material was sandwiched between two aluminium electrodes studded with holes to let moisture pass through. The resulting power illuminated a light-emitting diode lightbulb.

Energy Environ. Sci. <http://doi.org/bcg2> (2016)

TISSUE ENGINEERING

Polymers bolster printed tissue

A 3D printer can generate tissues using cells and polymers that are larger and more robust than previously printed biological structures.

Bioprinted 3D organs could one day help people who need transplants, but existing methods tend to produce only small, simple structures. Larger ones lose their shape, or die because nutrients cannot reach their centres. To build larger, stronger tissues, Anthony Atala and his colleagues at Wake Forest Institute for Regenerative Medicine in Winston-Salem, North Carolina, devised a 3D-printing system that adds biodegradable polymers for structural support. By combining polymer-based frames with hydrogels containing cells, the researchers printed a human-sized ear (**pictured**), a human jawbone fragment, a segment of mouse muscle and a piece of rat skull. Microchannels printed in the structures helped nutrients to flow into the tissues.

PROC. NATL ACADE SCI USA



The team implanted some of the structures into rodents and found that the tissues survived over weeks and months.

Nature Biotech. <http://dx.doi.org/10.1038/nbt.3413> (2016)

HYDROLOGY

Stored water slows rising seas

Changes in water storage on land may have slowed sea-level rise during the past decade.

John Reager of NASA's Jet Propulsion Laboratory in Pasadena, California, and his team investigated the shifting volumes of water stored on land using global data from NASA's Gravity Recovery and Climate Experiment (GRACE) satellite, which calculates water and ice mass on the basis of changes in Earth's gravity field. They found that between 2002 and 2014, 3,200 gigatons more water than expected was stored on land as snow, soil moisture, surface water and groundwater, thanks to climate-driven changes in hydrology. This offset sea-level rise caused by melting glaciers and ice sheets by about 20% over the same period.

These results show that climate-driven land water storage is significant enough to be included in future estimates of sea-level rise, the authors say.

Science 351, 699–703 (2016)

CANCER BIOLOGY

Metabolism varies within tumours

Human lung-tumour cells break down sugars in different ways in different patients and even in the same tumour.

Cells in the same tumour are known to vary genetically. To study tumour metabolism, Ralph DeBerardinis at the University of Texas Southwestern Medical Center in Dallas and his colleagues infused a harmless carbon isotope into nine people who had lung cancer, and combined clinical-imaging techniques

SOCIAL SELECTION

Popular topics on social media

Do female programmers face bias?

Female software developers see their contributions of code accepted more frequently by the open-source software repository GitHub than do men, according to a preprint that attracted much attention last week on social media. But this happened only when the contributor's gender was not obvious from their GitHub profile page. When gender was made clear, the acceptance rate for women fell to slightly less than that for men, say researchers who analysed data on the activity of more than 1.4 million users of GitHub. Morgan Ernest, an ecologist at the University of Florida in Gainesville, tweeted a link to the paper: "Well that's horrifying. Is that suggested programming change by a woman? Reject. Don't know it's a woman? Accept." But

► NATURE.COM

For more on popular papers: go.nature.com/qgflfl

other researchers questioned whether the study, posted at PeerJ PrePrints on 9 February, definitively demonstrates gender bias.

Preprint at <http://doi.org/bchz> (2016)



with mass spectrometry to track how the carbon was biochemically processed by the tumours. Cancer cells are thought to feed on glucose and release a by-product called lactate, but the team found that some tumour cells consumed both lactate and glucose. Blood-flow patterns showed that different parts of a single tumour had varying metabolic patterns.

Understanding these patterns could help to improve metabolic therapy for cancer, the authors suggest.

Cell 164, 681–694 (2016)

BIOMECHANICS

Cockroaches inspire robot

Researchers have discovered how cockroaches can speed through gaps just millimetres high — and have used their findings to build a compressible robot.

Kaushik Jayaram and Robert Full at the University of

California, Berkeley, observed American cockroaches (*Periplaneta americana*) as they squeezed through a series of crevices that decreased in height. The insects could maintain speeds of up to 60 centimetres per second after entering these tight areas, and only slowed when the ceiling height reached 4 millimetres — about one-third of the insects' free-standing height. They achieved their speed by using their legs and feet to push against friction between their bodies, the ceiling and the ground.

The cockroaches' mode of locomotion inspired the development of a soft-bodied robot (**pictured left**) that can compress its height by half (**pictured right**), allowing it to move through a tight space.

Proc. Natl Acad. Sci. USA <http://doi.org/bch5> (2016)

► NATURE.COM

For the latest research published by *Nature* visit: www.nature.com/latestresearch

SEVEN DAYS

The news in brief

FUNDING

Obama's budget

US President Barack Obama's fiscal year 2017 budget plan, released on 9 February, calls for a 4% bump in research and development funding. But science advocates and lawmakers are unhappy with Obama's decision to boost science by relying on 'mandatory' spending rather than on the 'discretionary' funding stipulated by Congress, which decides how much money each agency will receive. Obama has requested \$8 billion for the National Science Foundation, up 6.7% over 2016's estimate. The request for the National Institutes of Health is \$33.1 billion, up 2.6%, and NASA's budget request is \$19 billion, down \$300 million. See go.nature.com/hr9j6d for more.

RESEARCH

Resistance rising

Bacteria in humans and animals across Europe continue to show significant resistance to common antimicrobial treatments. The European Centre for Disease



CHRIS MADDALONI/NATURE

Gravitational waves ring out

Delighted physicists at the Advanced Laser Interferometer Gravitational-Wave Observatory (LIGO) revealed on 11 February that they had heard the gravitational 'ringing' produced by the collision of two black holes about 400 million parsecs (1.3 billion light-years) from Earth. The black holes, about

36 and 29 solar masses, merged into a single, gravitational sink in space-time weighing 62 solar masses. The merger temporarily radiated more energy — in the form of gravitational waves — than all the stars in the observable Universe emitted as light in the same amount of time. See page 261 for more.

NUMBER CRUNCH

1,188

The number of species of living sharks, skates, rays and chimaeras, according to a list collated at the University of Hamburg, Germany. The list includes 509 species of shark, 630 skate and ray species and 49 species of chimaera fish, a cartilaginous species. Source: S. Weigmann *J. Fish Biol.* <http://doi.org/bcj4>; 2016.

Prevention and Control and the European Food Standards Agency said in a joint report on 11 February that "high to extremely high" levels of resistance to multiple drugs were seen in some types of *Salmonella* in 2014. Resistance that would "significantly" reduce the effective treatments available was also seen in some *Campylobacter* types. Resistance to colistin — a last-resort antibiotic treatment — in *Salmonella* and *Escherichia coli* was found in chickens.

Research standards
A global network of science academies has called for coherent international research standards. The InterAcademy Partnership issued *Doing Global Science* (see go.nature.com/3zpmwg).

a 160-page guide to responsible science practices, on 11 February. The report says that researchers have an obligation to share data, conduct peer review, disclose conflicts of interest, keep clear records, and discuss possible harmful consequences of work when they plan projects. It also says that institutions should provide clear guides and strong training for responsible research.

Asian genomes

An effort to sequence the genomes of 100,000 people from Asia was announced on 11 February. The GenomeAsia 100K project, hosted by Nanyang Technological University in Singapore, aims to characterize the genomes of individuals from an initial

list of 19 countries across the continent. The project will also gather medical information and microbiome samples. The project consortium says that the biology of the world's Asian population is under-represented in scientific research, and hopes that its results will accelerate precision-medicine applications for Asian patients.

POLICY

Renewables rise

Energy from renewable sources reached 16% in 2014 in the European Union, up from 15% in 2013, according to 10 February figures from Eurostat, the EU's statistical office. The EU as a whole has a target to get 20% of its energy from renewables by 2020,

REUTERS/STEFFI LOOS with individual countries having different burdens. In 2014, Finland had the biggest increase in renewables, increasing by 2 percentage points, whereas Bulgaria slipped by 1 percentage point — although both countries remain on track to make their targets by 2020.

PEOPLE

Italy GM reprimand

The University of Naples in Italy formally reprimanded 11 scientists, including laboratory chief Federico Infascelli, on 10 February, after an investigation committee concluded that three research papers co-authored by them contained manipulated data. The papers' two corresponding authors are also forbidden from publishing for two years without approval from their department director. The papers claimed that animals were harmed by eating genetically modified feed. Police in the Veneto region used Infascelli's work to justify destruction of a genetically modified crop last year; the farmer is now seeking compensation.

Pachauri protests

Rajendra Pachauri (pictured) was appointed executive vice-chairman of Delhi-based think tank The Energy and



Resources Institute (TERI) on 8 February. Last year, Pachauri resigned as chair of the Intergovernmental Panel on Climate Change after claims that he had sexually harassed a now-former TERI employee. He was also replaced as TERI director-general. On 10 February, a second former employee accused Pachauri of sexual harassment. Alumni of TERI University have started a petition in protest against Pachauri's appointment, and the scientist has taken a voluntary leave of absence.

KI head resigns

Anders Hamsten, the vice-chancellor of the Karolinska Institute (KI) in Stockholm, has resigned after acknowledging that he mishandled an investigation into surgeon Paolo Macchiarini. The KI will also reopen its misconduct inquiry into Macchiarini, it announced on 13 February. In an open letter in the

Swedish newspaper *Dagens Nyheter*, Hamsten wrote that he had "completely misjudged" Macchiarini, who has been accused of research misconduct and ethical breaches in his work implanting artificial windpipes into patients. Last August, Hamsten cleared Macchiarini after an external investigation found that the surgeon had committed misconduct in seven research papers. See go.nature.com/7bpza6 for more.

EVENTS

Gas leak plugged

A huge gas leak at a storage facility of the Southern California Gas Company has been "temporarily controlled", the firm says. A 13 February assessment by California's Air Resources Board suggests that 94,000 tonnes of natural gas have escaped during the leak in Aliso Canyon, which was first detected on 23 October, and declared an emergency by the state governor in January. In an 15 February statement, Southern California Gas Company said that it was now cementing the leak to permanently seal the well.

LISA masses freed

Less than a week after the first detection of gravitational waves was announced (see page 261), a European

SEVEN DAYS THIS WEEK

COMING UP

21–26 FEBRUARY

New Orleans in Louisiana plays host to the 2016 Ocean Sciences Meeting.

osm.agu.org/2016

22–28 FEBRUARY

The 4th session of UNESCO's Plenary of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) convenes in Kuala Lumpur.

go.nature.com/p43aaz

Space Agency (ESA) probe has demonstrated a crucial technology for future space-based detectors of the waves. On 16 February ESA said that its LISA Pathfinder mission performed one of its most delicate stages — to suspend two small solid test cubes freely in a vacuum enclosure inside the probe. The spacecraft launched on 3 December and on 23 February will begin its science mission by bouncing a laser beam between the two cubes in freefall.

Philae's long sleep

The European Space Agency (ESA) has given up hope that their comet-probing lander, Philae, will ever work again. On 12 February, the space agency announced that Philae, the Rosetta spacecraft's lander, was entering "eternal hibernation". Philae landed on comet 67P/Churyumov-Gerasimenko in November 2014 and operated for 64 hours before going into hibernation. The craft last sent a signal to scientists in July last year. ESA scientists are no longer sending signals to the lander, and say the chances of hearing from it again are "close to zero".

NATURE.COM

For daily news updates see:
www.nature.com/news

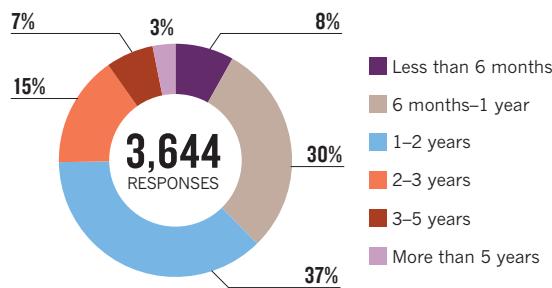
TREND WATCH

SOURCE: NATURE

An online poll answered by more than 3,600 *Nature* readers suggests that some 10% have waited at least 3 years for one or more of their papers to be published in a journal (see *Nature* 530, 148–151; 2016). But more than one-third had never waited longer than a year. Readers were also asked what they thought was the best way to speed up publication. More than 40% suggested that peer reviewers should stop asking for unnecessary revisions, and 22% asked journal editors to make quicker, clearer decisions.

THE WAITING GAME

Around 10% of *Nature*'s readers say that their longest wait to get a paper published in a journal has been 3 or more years.



Poll question: What is the longest time that you have waited for a research paper to be published?

NEWS IN FOCUS

LAB LIFE The environmental factors that ruin mouse experiments [p.264](#)

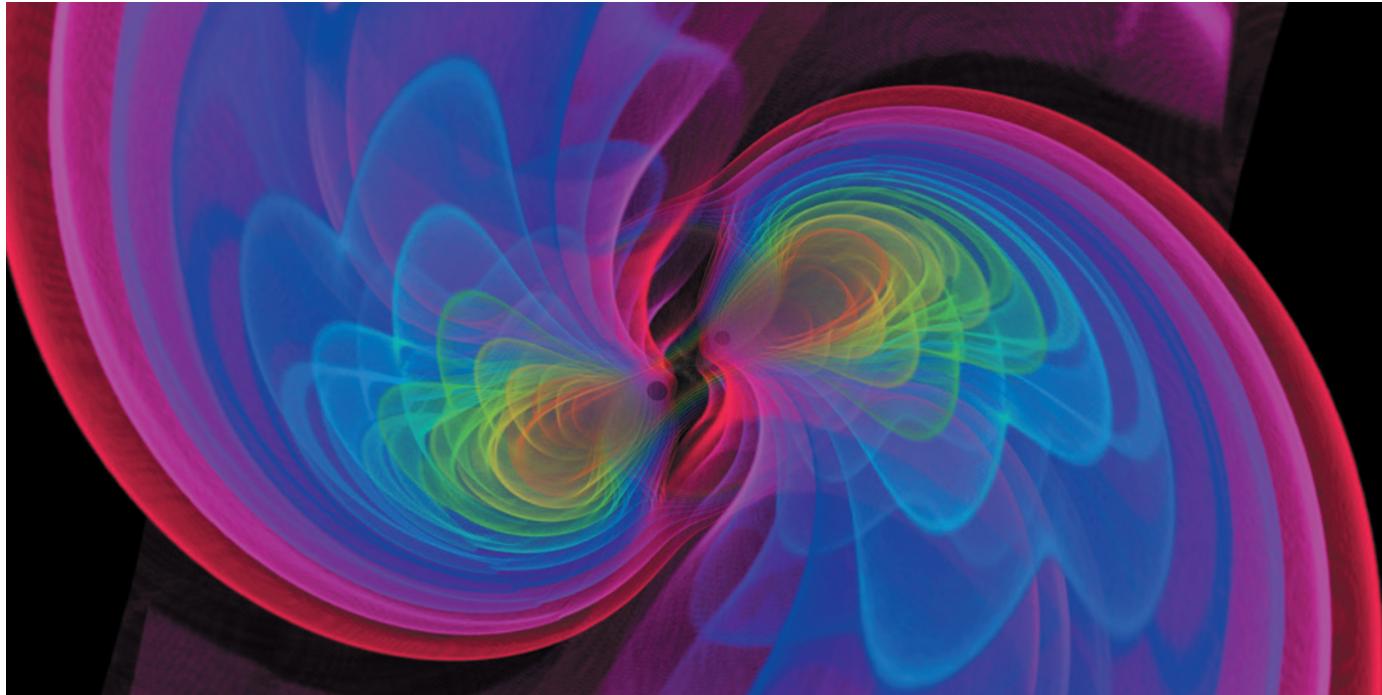
PUBLISHING Biologists urged to hug a preprint — and accelerate research [p.265](#)

POLLUTION Delhi driving ban yields trove of data [p.266](#)

EXOPLANETS The race to probe atmospheres of other worlds [p.272](#)



S. OSSOKINE, A. BUONINNO (MAX PLANCK INST. GRAVITATIONAL PHYS.), SCIENTIFIC VISUALIZATION; W. BENGEL (AIRBORNE HYDRO MAPPING)



The pair of merging black holes that LIGO detected using gravitational waves — as produced by a computer simulation.

GRAVITATIONAL WAVES

LIGO's path to victory

Historic discovery of ripples in space–time meant ruling out the possibility of a fake signal.

BY DAVIDE CASTELVECCHI

At 11:53 a.m. local time on 14 September 2015, an automated e-mail appeared in the inbox of Marco Drago, a physicist at the Max Planck Institute for Gravitational Physics in Hannover, Germany. It contained links to two plots, each showing a wave shaped like a bird's chirp that emerged suddenly from a noisy background and ended in a crash.

It was a signal that Drago had been trained to spot and that the US-led Advanced Laser Interferometer Gravitational-Wave Observatory (LIGO) that he works on was built to detect: the signature ripples in space–time produced when two black holes collide to form a single gravitational sink. No one had ever directly detected gravitational waves before, nor a black-hole merger. The plots, one from each of

LIGO's twin detectors in Washington state and Louisiana, would go on to make history.

On 11 February, the LIGO collaboration announced that it had made the first detection of gravitational waves from a black-hole merger that occurred about 400 million parsecs (1.3 billion light years) from Earth. It was just over 100 years after Albert Einstein predicted such waves as part of his general theory of relativity. "We did it!" David Reitze, the executive director of the LIGO Laboratory, said at a press conference in Washington DC.

As well as being expected to lead to a Nobel prize, LIGO's discovery launches the field of gravitational-wave astronomy, in which scientists will 'listen' to the waves to learn more

about the Universe (see page 263).

On that September morning, Drago could not take it for granted that he was looking at the chirp of a black-hole merger. "It was clear that it was something extraordinary," he says. But the plots were also something that the LIGO researchers had expected to see injected artificially by their colleagues to test the detectors. "I went down to the office of my colleague Andrew Lundgren to ask him if he was aware of an injection," says Drago.

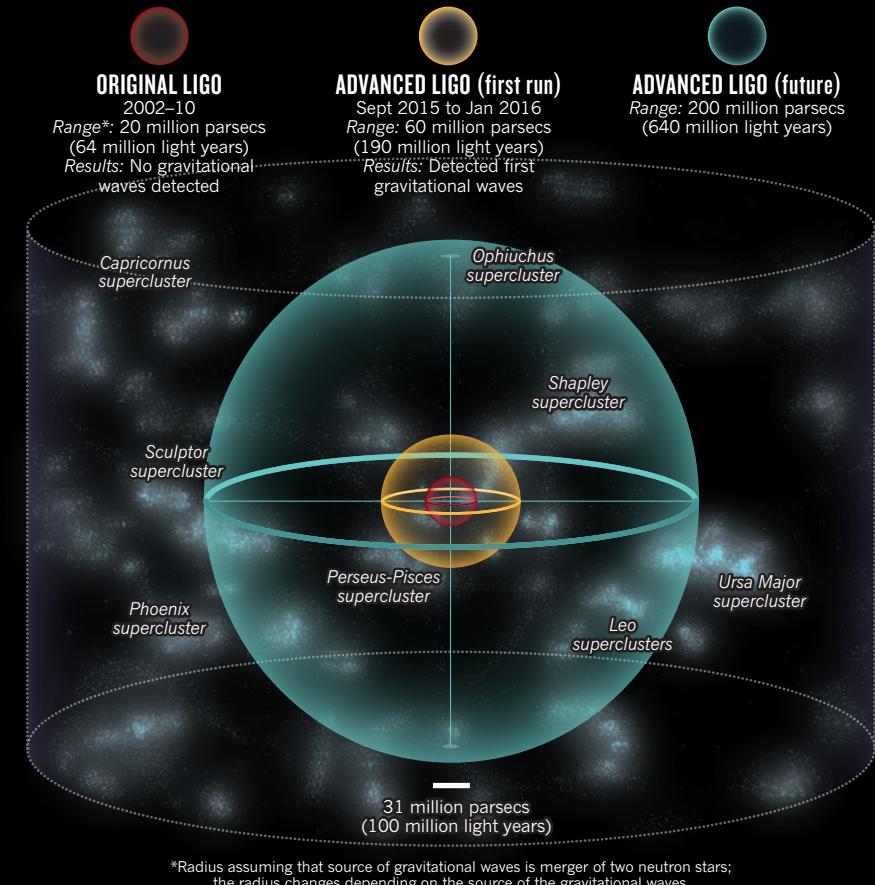
Lundgren quickly checked the data logs and found no traces of a drill. Next, Drago sent an e-mail to the entire LIGO collaboration — 1,000 researchers spread around the world — to see what they thought.

"When I first saw it, I said 'Oh, it's an injection, obviously,'" says physicist Bruce Allen, Drago and Lundgren's boss. Allen, who was

© NATURE.COM
For more on
gravitational waves
and LIGO see:
go.nature.com/f5crzd

LIGO'S GROWING UNIVERSE

An upgrade to the Laser Interferometer Gravitational-Wave Observatory (LIGO) that was completed last year increased the volume of space that the detector could scan. This greatly improved its chances of detecting gravitational waves — a historic feat that LIGO announced on 11 February. Further planned improvements mean that LIGO's observable universe is set to increase again soon.



To extract as much information as possible, the researchers then performed lengthy supercomputer simulations, Allen says. These confirmed that the data beautifully matched the predictions of Einstein's general theory of relativity in 1915, and the theoretical work that in the past few decades has led physicists to understand the theory's implications in fine detail.

CLUSTER MAP ADAPTED FROM ANDREW Z. COLVIN/CC-BY-SA

From the waveforms, the researchers were able to deduce that one black hole was about 36 times the mass of the Sun, and the other was about 29 solar masses. As the two objects orbited each other, they warped the fabric of space and time around them in a fluctuating pattern. Those fluctuations then travelled across the Universe as gravitational waves for an estimated 1.3 billion years, stretching and squeezing space as they moved.

LIGO's twin interferometers bounce laser beams between mirrors at the opposite ends of perpendicular, 4-kilometre-long vacuum pipes. A gravitational wave passing through will alter the length of the pipes in different ways, causing the laser beams to shift slightly out of sync. By the time the waves from the black-hole merger arrived on 14 September, they had become tiny ripples, changing the length of the pipes on the order of just 1 part in 1 billion trillion (10^{21}).

Although the two black holes had probably been orbiting each other for millions of years, LIGO began to pick up their waves only when they reached a frequency of 35 cycles per second (hertz). The frequency rapidly increased to 250 hertz. The signal became chaotic and then rapidly died down; the whole thing was over within a quarter of a second. Crucially, both detectors saw it at roughly the same time — Livingston, in Louisiana, first and Hanford, in Washington, 7 milliseconds later. The delay is an indication of how the waves swept through Earth.

Then came writing the paper. This involved getting 1,000 researchers to agree on every detail, and took some 5,000 e-mails, says LIGO's chief detector scientist Peter Fritschel at MIT. On 21 January, the team submitted the paper, which *Physical Review Letters* published on 11 February (B. P. Abbott *et al.* *Phys. Rev. Lett.* **116**, 061102; 2016), the same day that LIGO held multiple press conferences around the world.

"These amazing observations are the confirmation of a lot of theoretical work, including Einstein's general theory of relativity, which predicts gravitational waves," says physicist Stephen Hawking at the University of Cambridge, UK.

LIGO's triumph is a fitting end to the tale that Einstein began. He never believed that black holes existed. But although astronomers had accumulated compelling evidence for black holes by observing their surroundings, notes Thibault Damour, a theoretical physicist at the Institute of Advanced Scientific Studies near Paris, the LIGO signal is "the first real direct proof of their existence". ■

► in a meeting at the time, did not bother to enquire until after his lunch break.

Within a few hours, collaborators on the other side of the Atlantic woke up to Drago's e-mail — including experimental physicist Rainer Weiss at the Massachusetts Institute of Technology (MIT) in Cambridge, who is credited as the chief inventor of LIGO. "When I started looking at these waveforms, they were something spectacular," he says.

To many, the timing of the signal seemed too good to be true: the collaboration had completed a five-year upgrade to its instruments (see 'LIGO's growing universe'). Moreover, the LIGO collaboration had also given a small number of its members the power to inject fake signals and to hide whether they were real or simulated in order to test the team's responses. But even such a 'blind injection' ought to have left some traces in the data, says LIGO spokeswoman Gabriela González, a physicist at Louisiana State University in Baton Rouge.

After a long day of calls and e-mails, she determined that no blind injection had occurred and told the entire collaboration.

Only then did Kip Thorne, a theoretical physicist at the California Institute of Technology (Caltech) in Pasadena who co-founded LIGO with Weiss and Caltech colleague Ronald Drever, realize that a 40-year-old dream had come true. But it was not yet time to pop the champagne. The collaboration needed to do more before announcing a discovery to the world. "That night at home, I celebrated by just smiling to myself, because I could not tell my wife yet," Thorne says.

González and her team decided to take data for another

month before beginning a full analysis: the researchers needed to record the natural noise present in their detectors to have something to compare with the chirp. They concluded that the odds of noise producing that loud pattern — and the very same pattern in both Louisiana and Washington at about the same time — were so low that it should only occur by chance less than once every 203,000 years.

"I celebrated by just smiling to myself, because I could not tell my wife yet."

ASTROPHYSICS

Young scientists poised to ride the gravitational wave

Detection of ripples in space-time kicks off new era in physics.

BY ALEXANDRA WITZE

The first direct detection of gravitational waves has opened a new window in physics and astronomy — rewarding a cohort of young researchers who gambled on finding evidence of a phenomenon that had long eluded physicists.

Conceived in the 1970s and built in the 1990s, the Laser Interferometer Gravitational-Wave Observatory (LIGO) has been promising results for decades. On 11 February, it finally delivered, when project scientists reported finding signals of the space-time ripples known as gravitational waves.

Its observations are now poised to reshape ideas about high-gravity environments — such as colliding black holes, exploding stars and the earliest moments of the Universe.

“The big game-changer for me is, we really do have data and we can finally test our theories,” says Samaya Nissanke, an astrophysicist at Radboud University in Nijmegen, the Netherlands.

Nissanke is among the legions of early-career researchers who got into gravitational physics hoping to use data from LIGO and similar detectors. “You have this unique probe into extreme gravity and extreme space-time in a way that really holds your imagination,” she says. “I was hooked.”

LIGO’s first phase ran for years without detecting any gravitational waves, but after a major upgrade in September last year, it took just days to find a signal. That strengthens the belief that it will catch many future waves. Physicists expect each additional discovery to bring fresh insight.

“The fact that this happened right away has really given us a boost,” says Laura Cadonati, a physicist at the Georgia Institute of Technology in Atlanta who oversees LIGO data analysis. But the field did not look nearly so rosy 15 years ago, when Vicky Kalogera, an astrophysicist at Northwestern University in Evanston, Illinois, was calculating how often astrophysical objects such as black holes or neutron stars — the ultra-dense leftovers of exploded stars — merge. Such collisions are thought to be the source of most of the gravitational waves that LIGO was designed to detect.



Astrophysicist Mansi Kasliwal hopes to use gravitational-wave signals to study colliding neutron stars.

Kalogera led some of the early calculations to explore how often two neutron stars might collide close enough to Earth for LIGO to spot the ensuing gravitational waves (V. Kalogera *et al. Astrophys. J.* **601**, L179–L182; 2004). Estimates from different groups varied widely, and included some pessimistic scenarios in which LIGO had little chance of ever catching any waves. Kalogera faced a tough decision: whether to stick with gravitational-wave astrophysics or switch to topics that might be more likely to yield actual data.

“I went with my guts when everybody told me it was the wrong career choice,” she says. “Now, it is stunning to actually be in the detection era.”

Surprisingly, LIGO’s first detection did not come from a binary neutron-star system — which are thought to be relatively common, with six known pairs in our Galaxy alone — but from two large black holes. Both were of the order of 30 times the mass of the Sun. “You can start to think of these not just as gravitational-wave sources,” says Nissanke. “They are real astronomical beasts.”

Still, many physicists hold out hope that LIGO and similar detectors will soon catch gravitational waves from merging neutron stars. These incredibly dense stars are impenetrable to ordinary astronomical telescopes, which cannot

probe beneath their blazingly bright surfaces; researchers must rely on models to extrapolate what is going on inside.

Gravitational waves could change that, yielding information such as the precise sizes of neutron stars and how neutrons pack themselves together so tightly. These answers would come from the details of how the neutron stars spiral towards one another in the last moments before their final merger.

“There’s this potential to learn about the densest stable matter in the Universe, in a way that we’ve been blind to before,” says Jocelyn Read, a physicist at California State University, Fullerton.

When neutron stars merge, they are thought to fuse light chemical elements into heavier ones, which they then spew into the surrounding environment. Such cosmic collisions are the source of many of the heavy metals in the cosmos, including much of the gold that has ever been created, says Mansi Kasliwal, an astrophysicist at the California Institute of Technology in Pasadena.

“We haven’t actually seen explosions that are powerful enough to synthesize these elements,” she says. But when LIGO detects gravitational waves, astronomers will be able to command their telescopes to sweep the part of sky where the waves come from — and, with any luck, will capture a flash of these gold mines in the sky.

Kasliwal is already searching with a wide-field camera on a 1.2-metre telescope at Palomar Observatory in California. Next year, this effort will upgrade to a much bigger camera that can survey the sky 12 times faster. A similar survey in Chile is also expected to come online next year, giving both the Northern and Southern hemispheres a dedicated telescope for following the tantalizing traces of gravitational-wave detections.

Astronomers hope eventually to piece together a more complete picture of how gravitational waves and conventional astronomy come together. It’s like seeing a film in which the combination of images (electromagnetic waves) and sound (gravitational waves) provides a much fuller picture than either could alone, says Alessandra Corsi, an astrophysicist at Texas Tech University in Lubbock. “It feels incredibly exciting to be right at the start of a new era.” ■

LANCE HAYASHIDA/CALTECH



Mice are sensitive to minor changes in food, bedding and light exposure.

REPRODUCIBILITY

A mouse's house may ruin studies

Environmental factors lie behind many irreproducible rodent experiments.

BY SARA REARDON

It's no secret that therapies that look promising in mice rarely work in people. But, too often, experimental treatments that succeed in one mouse population do not even work in other mice, suggesting that many rodent studies are flawed from the start.

"We say mice are simpler, but I think the problem is deeper than that," says Caroline Zeiss, a veterinary neuropathologist at Yale University in New Haven, Connecticut. Researchers rarely report on subtle environmental factors such as their rodents' food, bedding or exposure to light; as a result, conditions vary widely across labs despite research showing that these factors can significantly affect the animals' biology.

"It's sort of surprising how many people are surprised by the extent of the variation" between mice that receive different care, says Cory Brayton, a pathologist at Johns Hopkins University in Baltimore, Maryland. At a meeting on mouse models at the Wellcome Genome Campus in Hinxton, UK, on 9–11 February, she and others explored the many biological factors that prevent mouse studies from being reproduced.

Christopher Colwell, a neuroscientist at the

University of California, Los Angeles, has first-hand experience with these issues. He and a colleague studied autism in the same genetically modified mouse line, but got different results on the same behavioural tests. Eventually, they worked out why: Colwell, who studies circadian rhythms, keeps his mice dark in the daytime to trick their body clocks into thinking day is night, so that the nocturnal animals are alert when tested in the day. His colleague does not.

Nutrition can also determine whether a mouse study succeeds or fails. Some mouse foods contain oestrogens and endocrine-disrupting chemicals that can affect research on cancer, among other diseases (M. Nygaard Jensen and M. Ritskes-Hoitinga *Lab. Anim.* **41**, 1–18; 2007). And the high-fat, high-sugar food used in obesity studies goes rancid quickly; when it does, mice may stop eating and lose weight without researchers realizing why.

Food choices can also alter a mouse's gut microbiome. Catherine Hagan Gillespie, a veterinary pathologist at the Jackson Laboratory in Sacramento, California, has found that species of bacteria in the gut vary widely between mice from different vendors (A. C. Ericsson *et al. PLoS ONE* **10**, e0116704; 2015). In an

unpublished study, she also found that mice with different gut bacteria showed different anxiety levels in behavioural tests.

But few behavioural scientists think about microbiology assessments, says Hagan Gillespie. Even when they do, the extra work can increase the complexity and cost of the study. Yet the mouse microbiome is sensitive to many factors, such as air quality and maternal stress.

Differences in the gut microbiome may explain why mice with the same genes can have different characteristics, or phenotypes, says George Weinstock, associate director for microbial genomics at the Jackson Lab's site in Farmington, Connecticut. The lab, which breeds and supplies mice for use in studies around the world, tightly controls factors such as the type and quantity of food and the pH of water that animals receive. Even so, it finds differences between mice at its three sites. Weinstock says that the company has begun looking into standardizing its customers' experiments by providing food and care instructions for its mice.

But even when improved mice and food are available, some researchers resist using them in case it affects their results, says Graham Tobin, former technical director of mouse-diet vendor Teklad in Alconbury, UK. He argues that standardizing results is worth any inconvenience, and notes that researchers rarely resist adopting other new technologies that can throw older data into question.

Zeiss says that the competitive nature of science might increase researchers' resistance to changing how they consider animals in research design. If scientists have to treat their animals at the right point in the experiment, analyse both clinical and biomarker changes, include old mice and both sexes to ensure that results are representative of broad populations, and control environmental variables, each experiment will take much longer and the scientists are probably not going to be able to publish as much, she says.

The US National Institutes of Health (NIH) has taken steps to address some of these problems, although some people say it is not enough. Some NIH institutes require certain animal trials to be replicated before a therapy can move into clinical trials, but the agency says that it has no plans to require this agency-wide. And in 2014, the NIH began requiring researchers to include female animals in studies, and giving out supplementary grants to those who complained about the cost. But the agency has not issued any specific grants or supplements to study other confounding factors.

That is disappointing to those who would like to see researchers control — or at least report — factors such as the strain of mice used and the type of environment they are raised in. This would allow meta-analyses of published literature that could identify any confounding factors. "The information and the wisdom is out there," says Zeiss, "but studies get funded without necessarily a lot of attention to that." ■ **SEE EDITORIAL P.254**

PUBLISHING

'Hug a preprint, biologists!'

ASAPbio meeting discusses the ins and outs of posting work online before peer review.

BY EWEN CALLAWAY AND KENDALL POWELL

Physicists do it; computer scientists, mathematicians and economists do it.

And this week, a who's who of biomedical researchers and publishers is asking what it will take to convince life scientists to do it, too — release their work online before peer review and formal journal publication.

The impetus for the gathering, called ASAPbio (asapbio.org), is the growing frustration of some researchers at the slow pace of publishing in biology journals (see *Nature* 530, 148–151; 2016). The delay can take years, notes Ron Vale, a cell biologist at the University of California, San Francisco. That can seriously affect scientists' careers because they don't receive recognition for their work until it is published.

The solution, argues Vale, a co-organizer of ASAPbio, is for biologists to embrace preprints: pre-publication manuscripts posted online. These speed up dissemination, give students and postdocs tangible ways to cite their contributions to the literature, and stimulate discussion and ideas, he says — accelerating and improving life-sciences research.

There are signs that some biologists are ready to follow the lead of their colleagues in the physical sciences, where it is now routine for research to be submitted to the arXiv preprint server — founded 25 years ago — before publication. A life-sciences-only preprint server called bioRxiv started in 2013 and is rapidly growing in popularity (see 'The growth of bioRxiv'), especially in data-intensive fields such as computational biology and genomics.

It has now seen more than 3,100 posted preprints, says John Inglis, the site's co-founder and the executive editor of Cold Spring Harbor Laboratory Press in New York. Other journals, such as the online *F1000Research*, also encourage the posting of life-sciences manuscripts before peer review.

But preprints are still unfamiliar ground for biologists, Vale says. Leslie Vosshall, a neurobiologist at the Rockefeller University in New York City, says that if such sites are to become popular in the life sciences, researchers will have to overcome common concerns — for example, that preprints could lead to scientists being scooped by competitors and missing out on credit for ideas. Vale and Vosshall say that such worries are misplaced. "I think most biologists don't know about preprints, or if they do, they've heard of them at a very superficial level, to the point that they don't really understand them very well," Vale says.

life-sciences community develops a consensus that preprint publication establishes a priority for any discovery. A discussion about that is at the top of ASAPbio's agenda, and Vale co-authored an article on it, posted to the conference's website last week. He has also tasked meeting attendees with considering how funding agencies and academic committees should view preprints when deciding whom to fund and hire.

Another concern is that quality might dip if life scientists flood preprint servers with non-peer-reviewed work. But supporters of preprint publication say that, if anything, researchers are more careful when their reputation rides on early work made public for all to critique.

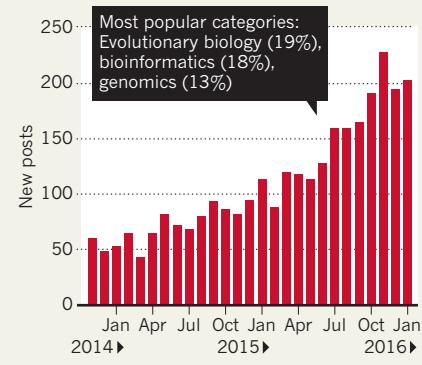
The issue of whether a preprint could jeopardize the chances of a manuscript subsequently appearing in a peer-reviewed journal is also being resolved, says Inglis. Since bioRxiv launched, several journal publishers have changed their policies to expressly allow the publication of research previously posted to preprint servers.

Some scientists would like to see more-radical changes. Many make new data sets and hypotheses instantly and freely available online at repositories such as GitHub, figshare and Zenodo, and hope for crowdsourced peer review of their work. "That's my utopian fantasy. It would be amazing to live in a world with all radically free data," says Jessica Polka, a postdoctoral fellow at Harvard Medical School in Boston, Massachusetts, and co-organizer of the ASAPbio meeting. She says that preprints are "the most practical of all the transformative things that could be implemented".

All of Vosshall's preprint articles have also been published in conventional journals "through the excruciatingly slow process of peer review", she notes. "Most of them don't look any different. Which begs the question, why do we need journals any more?" ■

THE GROWTH OF BIORXIV

More than 3,100 preprints have been posted to the biology preprint server bioRxiv since its launch in 2013.



"There's no doubt that preprints are happening," says Harold Varmus, a cancer biologist at Weill Cornell Medical College in New York City and another co-organizer of ASAPbio, held on 16–17 February at the Howard Hughes Medical Institute in Chevy Chase, Maryland. "But I don't think we've ever had a conversation among all the constituents about what the effects will be."

Both Vale and Vosshall think that preprints will become widely accepted only if the

MORE ONLINE

Q & A



Howard Hughes's next president: "Promote under-represented groups in science" go.nature.com/erbvjg

MORE NEWS

- 'Water scarcity' affects four billion people each year go.nature.com/gwq3zp
- Karolinska's vice-chancellor resigns go.nature.com/7bpz46
- Parkinson's patients trained to respond to placebos go.nature.com/qmtpfh

NATURE PODCAST



Making shipping greener; highlights from the AAAS conference; and human genes in a Neanderthal nature.com/nature/podcast

CLIMATE CHANGE

UN sets out emissions plan for planes

Standards aim to reduce CO₂ produced by new aircraft.

BY JEFF TOLLEFSON

A United Nations panel has proposed the first global greenhouse-gas emissions standard for aircraft.

The draft rule, released by the International Civil Aviation Organization (ICAO) on 8 February, would apply to most new commercial and business aircraft, including designs already in production. But it would require only minimal changes to aviation technology over the next 12 years, and many environmentalists say that the proposal is inadequate.

The ICAO standard would take full effect in 2028; the panel is expected to adopt the plan later this year. Doing so could cut the amount of fuel used at cruising speed by an average of 4% compared with current levels, according to the International Council on Clean Transportation (ICCT), a non-profit research group in Washington DC.

But many environmental groups found the proposal wanting. "We think that this is just woefully insufficient," says Vera Pardee, a lawyer with the Center for Biological Diversity in Oakland, California. She notes that the plan would not apply to aircraft that are already flying.

Daniel Rutherford, programme director for marine and aviation at the ICCT in San Francisco, California, agrees that the ICAO could have been more aggressive. An ICCT study released in December found that manufacturers could reduce fuel consumption in new planes by 25% in 2024 and by 40% in 2034 by improving engine technologies and aerodynamics, and reducing aircraft weight.

Nonetheless, Rutherford says, the ICAO's plan is a step forward. "These standards do tend to matter over time as you update them and make them more stringent," he adds.

The ICAO process aims to plug a gap in the UN climate agreement signed in Paris last December. The panel also is working on a market-based offset mechanism that would levy a fee on international flights.

In the meantime, individual countries might implement more-stringent standards for aircraft emissions, and environmentalists are gearing up for a fight. For instance, lawsuits from environmental groups helped to push the US Environmental Protection Agency to begin developing its own greenhouse-gas standards for aircraft. ■



Pollution in New Delhi fell by 10% when vehicle numbers on its roads were temporarily reduced.

ATMOSPHERIC SCIENCE

Car ban yields science bounty

Scramble by researchers to monitor driving restrictions in Indian capital pays off.

BY MEERA SUBRAMANIAN

New Delhi may be the world's most polluted city, but it's making an effort to relinquish that title. With pollution from particulate matter at potentially lethal levels early last December, city officials took a drastic step: they announced that they would temporarily restrict the use of private vehicles by allowing owners to drive only on alternate days, based on the sequence of their number plates.

The initial results of that 15-day trial, which began on 1 January, are now in. Although traffic actually increased in the first week of the ban, the levels of PM_{2.5} — particulate matter measuring less than 2.5 micrometres across — fell by roughly 10%. That is a victory not just for New Delhi officials, but also for the scientists who sprang into action to collect the data necessary to determine whether the test had achieved its goal.

"This experiment with 'live research' has

been really quite exciting," says Santosh Harish, assistant director of the India centre of the Energy Policy Institute at the University of Chicago (EPIC-India). EPIC-India and the New Delhi-based Council on Energy, Environment and Water (CEEW), an independent think tank, used video monitors around the city to document the types and numbers of vehicles on the roads. The groups had less than a month to collect baseline data before the driving restrictions began.

But they weren't the only researchers interested in Delhi's living lab. Economist Gabriel Kreindler of the Massachusetts Institute of Technology in Cambridge scrambled to secure human-study approval and funding for a survey of driver behaviour during the traffic restrictions. Within 18 days of the announcement of the driving ban, he had arrived in New Delhi to oversee a surveying team from the Abdul Latif Jameel Poverty Action Lab's office there. Kreindler's work eventually found that the alternate-day restrictions were

MEERA SUBRAMANIAN

well received by most drivers, who, in spite of the disruption, were willing to comply and alter their behaviour for short periods of time.

Other researchers built on work already under way. The Centre for Science and Environment (CSE), a non-profit research and advocacy group in New Delhi, had been closely analysing government air-quality data since last October. By December, government monitors were recording daily levels of noxious PM_{2.5} in the range of 400–600 micrograms per cubic metre. This is much higher than the Indian legal standard of 60 micrograms (which itself is more than double the 25-microgram target threshold set by the World Health Organization).

PM_{2.5} particles cause more than 600,000 premature deaths in India each year, from lung cancer, asthma, and cardiovascular and respiratory diseases. There is no known safe level for this pernicious pollutant.

The CSE's analysis found that, despite unfavourable weather conditions, the peak pollution during the driving scheme was lower than it would have been without the restrictions in place.

“The region is geographically disadvantaged,” says M. P. George, a scientist with the government's Delhi Pollution Control Committee. In winter, particulate levels can be

twice as high as during the summer, because ‘inversion layers’ of warm air trap cold air close to the ground. This prevents pollution from dissipating into the atmosphere. Emissions from

“This experiment with ‘live research’ has been really quite exciting.”

vehicles and construction dust also combine with raised levels of black carbon generated from winter sources — fires for warmth, brick kilns that are lit in the autumn, and widespread field burning in neighbouring states.

“It's a very simple math,” says Sarath Guttikunda, director of the independent research group Urban Emissions, which is registered in New Delhi. “In winter, your air volume is going down and your emissions are going up.”

Because atmospheric conditions such as wind and temperature can greatly affect particulate-matter measurements, researchers from EPIC-India and the Evidence for Policy Design initiative at Harvard University in Cambridge, Massachusetts, gathered data from air-quality monitors in New Delhi and placed monitors in three adjacent cities as a control. They found that the daily level of PM_{2.5} pollution in Delhi dropped by 10–13% during the vehicle restrictions. Hourly comparisons showed an even greater improvement, at times an 18% fall.

The question now is whether New Delhi, the capital of a nation with dozens of growing cities choked by pollution, can build on the experiment for long-term gains in air quality. “Delhi has to get it right,” says Namit Arora, a member of the pollution task force of the Delhi Dialogue Commission, a government initiative. This will require long-term strategies and coordination between local, regional and national efforts, he says, as well as a reduction in all sources of air pollution. Other researchers stress the need for more open-access data from a wide range of well-calibrated instruments.

But the driving-restriction experiment has given researchers a tantalizing glimpse of one possible future. “We need to re-imagine the way we think about cities,” says Hemanshu Dholakia, a research associate at the CEEW. “That's the real opportunity.” ■

CORRECTIONS

The Editorial ‘Blue future’ (Nature 529, 255–256; 2016) should have said that 2.4–4.6% of the world's carbon emissions are captured and sequestered by living organisms in the oceans. And the power of ASTRO-H probe is 3,500 watts not 2,500 as stated in the News story ‘High stakes for Japan's space probe’ (Nature 530,



JOHN SIBICK/NATURAL HISTORY MUSEUM

WHAT SPARKED THE CAMBRIAN EXPLOSION?

An evolutionary burst 540 million years ago filled the seas with an astonishing diversity of animals. The trigger for that revolution is finally coming into focus.

BY DOUGLAS FOX

A series of dark, craggy pinnacles rises 80 metres above the grassy plains of Namibia. The peaks call to mind something ancient — the burial mounds of past civilizations or the tips of vast pyramids buried by the ages.

The stone formations are indeed monuments of a faded empire, but not from anything hewn by human hands. They are pinnacle reefs, built by cyanobacteria on the shallow sea floor 543 million years ago, during what is known as the Ediacaran period. The ancient world occupied by these reefs was truly alien. The

oceans held so little oxygen that modern fish would have quickly foundered and died there. A gooey mat of microbes covered the sea floor, and on that blanket lived a variety of enigmatic animals whose bodies resembled thin, quilted pillows. Most were stationary, but a few meandered blindly over the slime, grazing on the microbes. Animal life was simple, and there were no predators. But an evolutionary storm would soon upend this quiet world.

Within several million years, this simple ecosystem would disappear, and give way to a world ruled by highly mobile animals that sported

modern anatomical features. The Cambrian explosion, as it is called, produced arthropods with legs and compound eyes, worms with feathery gills and swift predators that could crush prey in tooth-rimmed jaws. Biologists have argued for decades over what ignited this evolutionary burst. Some think that a steep rise in oxygen levels sparked the change, whereas others say that it sprang from the development of some key evolutionary innovation, such as vision. The precise cause has remained elusive, in part because so little is known about the physical and chemical environment at that time.

The Cambrian seas teemed with new types of animal, such as the predator *Anomalocaris* (centre).

But over the past several years, discoveries have begun to yield some tantalizing clues about the end of the Ediacaran. Evidence gathered from the Namibian reefs and other sites suggests that earlier theories were too simplistic — that the Cambrian explosion actually emerged out of a complex interplay between small environmental changes that triggered major evolutionary developments.

Some scientists now think that a small, perhaps temporary, increase in oxygen suddenly crossed an ecological threshold, enabling the emergence of predators. The rise of carnivory would have set off an evolutionary arms race that led to the burst of complex body types and behaviours that fill the oceans today. “This is the most significant event in Earth evolution,” says Guy Narbonne, a palaeobiologist at Queen’s University in Kingston, Canada. “The advent of pervasive carnivory, made possible by oxygenation, is likely to have been a major trigger.”

ENERGY TO BURN

In the modern world, it’s easy to forget that complex animals are relative newcomers to Earth. After life first emerged more than 3 billion years ago, single-celled organisms dominated the planet for most of its history. Thriving in environments that lacked oxygen, they relied on compounds such as carbon dioxide, sulfur-containing molecules or iron minerals that act as oxidizing agents to break down food. Much of Earth’s microbial biosphere still survives on these anaerobic pathways.

Animals, however, depend on oxygen — a much richer way to make a living. The process of metabolizing food in the presence of oxygen releases much more energy than most anaerobic pathways. Animals rely on this potent, controlled combustion to drive such energy-hungry innovations as muscles, nervous systems and the tools of defence and carnivory — mineralized shells, exoskeletons and teeth.

Given the importance of oxygen for animals, researchers suspected that a sudden increase in the amount of the gas in the oceans, to near-modern levels, could have spurred the Cambrian explosion. To test that idea, they have studied ancient ocean sediments laid down during the Ediacaran and Cambrian periods, which together ran from about 635 million to 485 million years ago.

In Namibia, China and other spots around the world, researchers have collected rocks that were once ancient sea beds, and analysed the amounts of iron, molybdenum and other metals in them. The metals’ solubility depends strongly on the amount of oxygen present, so the quantity and type of those metals in ancient sedimentary rocks reflect how much oxygen was in the water, when the sediments formed.

These proxies seemed to indicate that oxygen concentrations in the oceans rose in several

steps, approaching today’s sea-surface concentrations at the start of the Cambrian, around 541 million years ago — just before more-modern animals suddenly appeared and diversified. This supported the idea of oxygen as a key trigger for the evolutionary explosion.

But last year, a major study¹ of ancient sea-floor sediments challenged that view. Erik Sperling, a palaeontologist at Stanford University in California, compiled a database of 4,700 iron measurements taken from rocks around the world, spanning the Ediacaran and Cambrian periods. He and his colleagues did not find a statistically significant increase in the proportion of oxic to anoxic water at the boundary between the Ediacaran and the Cambrian.

“Any oxygenation event must have been far, far smaller than what people normally considered,” concludes Sperling. Most people assume “that the oxygenation event essentially raised oxygen to essentially modern-day levels. And that probably wasn’t the case”, he says.

The latest results come at a time when scientists are already reconsidering what was happening to ocean oxygen levels during this crucial period. Donald Canfield, a geobiol-

“This is the most significant event in Earth evolution.”

ogist at the University of Southern Denmark in Odense, doubts that oxygen was a limiting factor for early animals. In a study published last month², he and his colleagues suggest that oxygen levels were already high enough to support simple animals, such as sponges, hundreds of millions of years before they actually appeared. Cambrian animals would have needed more oxygen than early sponges, concedes Canfield. “But you don’t need an increase in oxygen across the Ediacaran–Cambrian boundary,” he says; oxygen could already have been abundant enough “for a long, long time before”.

“The role of oxygen in the origins of animals has been heavily debated,” says Timothy Lyons, a geobiologist at the University of California, Riverside. “In fact, it’s never been more debated than it is now.” Lyons sees a role for oxygen in evolutionary changes, but his own work³ with molybdenum and other trace metals suggests that the increases in oxygen just before the Cambrian were mostly temporary peaks that lasted a few million years and gradually stepped up (see ‘When life sped up’).

MODERN MIRRORS

Sperling has looked for insight into Ediacaran oceans by studying oxygen-depleted regions in modern seas worldwide. He suggests that biologists have conventionally taken the wrong approach to thinking about how oxygen shaped animal evolution. By pooling previously published data with some of his own and analysing

them, he found that tiny worms survive in areas of the sea floor where oxygen levels are very low — less than 0.5% of average global sea-surface concentrations. Food webs in these oxygen-poor environments are simple, and animals feed directly on microbes. In places where sea-floor oxygen levels are a bit higher — about 0.5–3% of concentrations at the sea surface — animals are more abundant but their food webs remain limited: the animals still feed on microbes rather than on each other. But around somewhere between 3% and 10%, predators emerge and start to consume other animals⁴.

The implications of this finding for evolution are profound, Sperling says. The modest oxygen rise that he thinks may have occurred just before the Cambrian would have been enough to trigger a big change. “If oxygen levels were 3% and they rose past that 10% threshold, that would have had a huge influence on early animal evolution,” he says. “There’s just so much in animal ecology, lifestyle and body size that seems to change so dramatically through those levels.”

The gradual emergence of predators, driven by a small rise in oxygen, would have meant trouble for Ediacaran animals that lacked obvious defences. “You’re looking at soft-bodied, mostly immobile forms that probably lived their lives by absorbing nutrients through their skin,” says Narbonne.

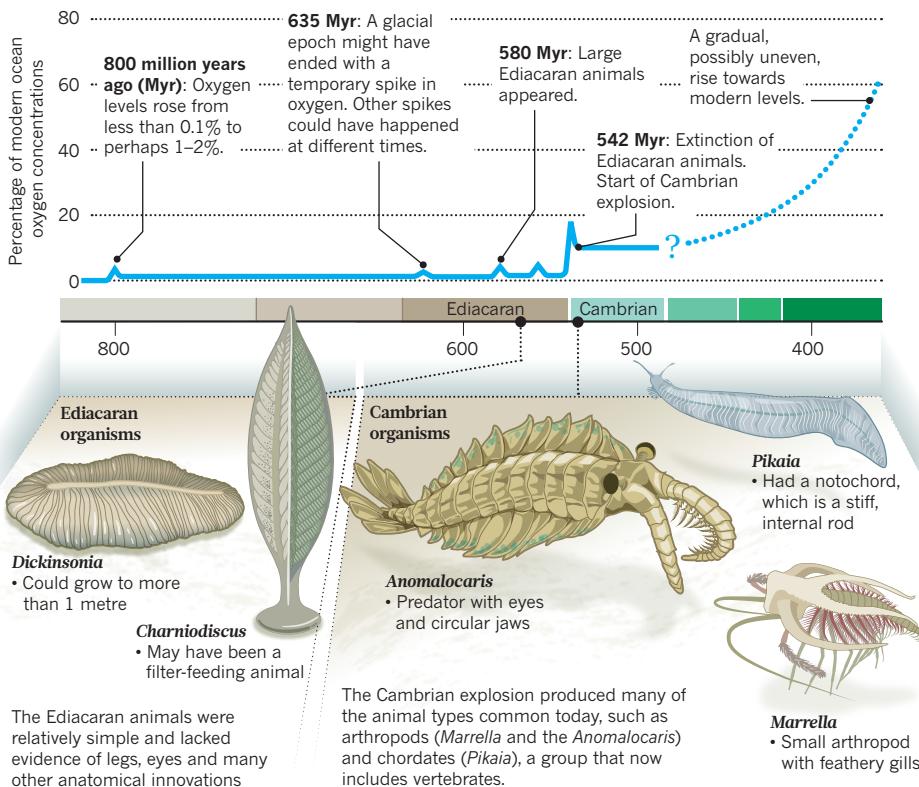
Studies of the ancient Namibian reefs suggest that animals were indeed starting to fall prey to predators by the end of the Ediacaran. When palaeobiologist Rachel Wood from the University of Edinburgh, UK, examined the rock formations, she found spots where a primitive animal called *Cloudina* had taken over parts of the microbial reef. Rather than spreading out over the ocean floor, these cone-shaped creatures lived in crowded colonies, which hid their vulnerable body parts from predators — an ecological dynamic that occurs in modern reefs⁵.

Cloudina were among the earliest animals known to have grown hard, mineralized exoskeletons. But they were not alone. Two other types of animal in those reefs also had mineralized parts, which suggests that multiple, unrelated groups evolved skeletal shells around the same time. “Skeletons are quite costly to produce,” says Wood. “It’s very difficult to come up with a reason other than defence for why an animal should bother to create a skeleton for itself.” Wood thinks that the skeletons provided protection against newly evolved predators. Some *Cloudina* fossils from that period even have holes in their sides, which scientists interpret as the marks of attackers that bored into the creatures’ shells⁶.

Palaeontologists have found other hints that animals had begun to eat each other by the late Ediacaran. In Namibia, Australia and Newfoundland in Canada, some sea-floor sediments have preserved an unusual type of tunnel made by an unknown, wormlike creature⁷. Called *Treptichnus* burrows, these warrens branch again and again, as if a predator just below the

When life sped up

Big animals emerged during the Ediacaran period, but these creatures were slow or immobile. A rise in oceanic oxygen concentrations at the end of the period might have helped to trigger the Cambrian evolutionary explosion.



The Ediacaran animals were relatively simple and lacked evidence of legs, eyes and many other anatomical innovations

microbial mat had systematically probed for prey animals on top. The *Treptichnus* burrows resemble those of modern priapulid, or 'penis', worms — voracious predators that hunt in a remarkably similar way on modern sea floors⁸.

The rise of predation at this time put large, sedentary Ediacaran animals at a big disadvantage. "Sitting around doing nothing becomes a liability," says Narbonne.

THE WORLD IN 3D

The moment of transition from the Ediacaran to the Cambrian world is recorded in a series of stone outcrops rounded by ancient glaciers on the south edge of Newfoundland. Below that boundary are impressions left by quilted Ediacaran animals, the last such fossils recorded on Earth. And just 1.2 metres above them, the grey siltstone holds trails of scratch marks, thought to have been made by animals with exoskeletons, walking on jointed legs — the earliest evidence of arthropods in Earth's history.

No one knows how much time passed in that intervening rock — maybe as little as a few centuries or millennia, says Narbonne. But during that short span, the soft-bodied, stationary Ediacaran fauna suddenly disappeared, driven to extinction by predators, he suggests.

Narbonne has closely studied the few fauna that survived this transition, and his findings suggest that some of them had acquired new, more complex types of behaviour. The best clues come from traces left by peaceful, wormlike

animals that grazed on the microbial mat. Early trails from about 555 million years ago meander and criss-cross haphazardly, indicating a poorly developed nervous system that was unable to sense or react to other grazers nearby — let alone predators. But at the end of the Ediacaran and into the early Cambrian, the trails become more sophisticated: creatures carved tighter turns and ploughed closely spaced, parallel lines through the sediments. In some cases, a curvy feeding trail abruptly transitions into a straight line, which Narbonne interprets as potential evidence of the grazer evading a predator⁹.

This change in grazing style may have contributed to the fragmentation of the microbial mat, which began early in the Cambrian. And the transformation of the sea floor, says Narbonne, "may have been the most profound change in the history of life on Earth"^{10,11}. The mat had previously covered the sea bed like a coating of plastic wrap, leaving the underlying sediments largely anoxic and off limits to animals. Because animals could not burrow deeply in the Ediacaran, he says, "the mat meant that life was two-dimensional". When grazing capabilities improved, animals penetrated the mat and made the sediments habitable for the first time, which opened up a 3D world.

Tracks from the early Cambrian show that animals started to burrow several centimetres into the sediments beneath the mat, which provided access to previously untapped nutrients — as well as a refuge from predators. It's

also possible that animals went in the opposite direction. Sperling says that the need to avoid predators (and pursue prey) may have driven animals into the water column above the sea bed, where enhanced oxygen levels enabled them to expend energy through swimming.

The emerging evidence about oxygen thresholds and ecology could also shed light on another major evolutionary question: when did animals originate? The first undisputed fossils of animals appear only 580 million years ago, but genetic evidence indicates that basic animal groups originated as far back as 700 million or 800 million years ago. According to Lyons, the solution may be that oxygen levels rose to perhaps 2% or 3% of modern levels around 800 million years ago. These concentrations could have sustained small, simple animals, just as they do today in the ocean's oxygen-poor zones. But animals with large bodies could not have evolved until oxygen levels climbed higher, in the Ediacaran.

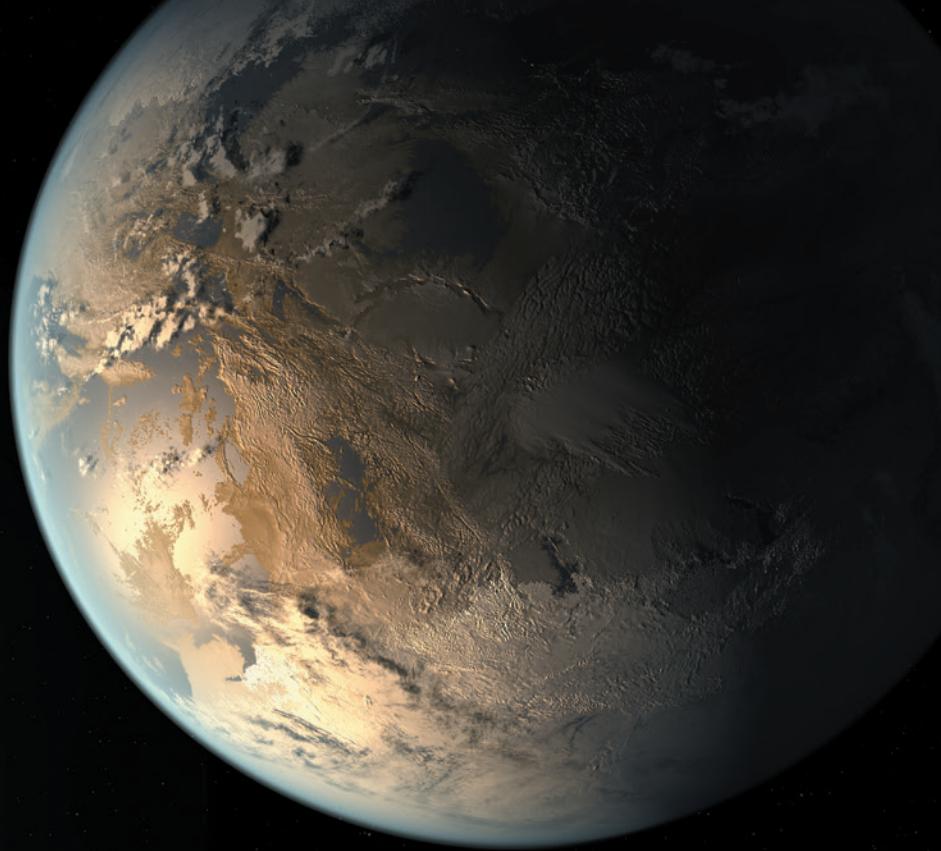
Understanding how oxygen influenced the appearance of complex animals will require scientists to tease subtler clues out of the rocks. "We've been challenging people working on fossils to tie their fossils more closely to our oxygen proxies," says Lyons. It will mean deciphering what oxygen levels were in different ancient environments, and connecting those values with the kinds of traits exhibited by the animal fossils found in the same locations.

Late last year, Woods visited Siberia with that goal in mind. She collected fossils of *Cloudina* and another skeletonized animal, *Suvorovella*, from the waning days of the Ediacaran. The sites gave her the chance to gather fossils from many different depths in the ancient ocean, from the more oxygen-rich surface waters to deeper zones. Woods plans to look for patterns in where animals grew tougher skeletons, whether they were under attack by predators and whether any of this had a clear link with oxygen levels. "Only then can you pick out the story." ■

Douglas Fox is a journalist in northern California.

1. Sperling, E. A. *et al.* *Nature* **523**, 451–454 (2015).
2. Zhang, S. *et al.* *Proc. Natl. Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1523449113> (2016).
3. Sahoo, S. K. *et al.* *Geobiology* (in the press).
4. Sperling, E. A. *et al.* *Proc. Natl. Acad. Sci. USA* **110**, 13446–13451 (2013).
5. Wood, R. A. *et al.* *Precambrian Res.* **261**, 252–271 (2015).
6. Bengtson, S. & Zhao, Y. *Science* **257**, 367–369 (1992).
7. Seilacher, A., Buatois, L. A. & Mángano, M. G. *Palaeogeog. Palaeoclimatol. Palaeoecol.* **227**, 323–356 (2005).
8. Vannier, J., Calandra, I., Gaillard, C. & Zylinska, A. *Geology* **38**, 711–714 (2010).
9. Carbone, C. & Narbonne, G. M. *J. Paleontol.* **88**, 309–330 (2014).
10. Mángano, M. G. & Buatois, L. A. *Proc. R. Soc. B* **281**, 20140038 (2014).
11. Buatois, L. A., Narbonne, G. M., Mángano, M. C., Carmona, N. B. & Myrow, P. *Nature Commun.* **5**, 3544 (2014).

Kepler-186f, the first known Earth-sized exoplanet in a star's habitable zone (artist's impression).



Astronomers are beginning to glimpse what exoplanets orbiting distant suns are actually like.

THE LIGHT OF OTHER WORLDS

BY JEFF HECHT

The trickle of discoveries has become a torrent.

Little more than two decades after the first planets were found orbiting other stars, improved instruments on the ground and in space have sent the count soaring: it is now past 2,000. The finds include 'hot Jupiters', 'super-Earths' and other bodies with no counterpart in our Solar System — and have forced astronomers to radically rethink their theories of how planetary systems form and evolve.

Yet discovery is just the beginning. Astronomers are aggressively moving into a crucial phase in exoplanet research: finding out what these worlds are like. Most exoplanet-finding techniques reveal very little apart from the planet's mass, size and orbit. But is it rocky like Earth or a gas giant like Jupiter? Is it blisteringly hot or in deep-freeze? What is its atmosphere made of? And does that atmosphere contain molecules such as water, methane and oxygen in odd, unstable proportions that might be a signature of life?

The only reliable tool that astronomers can use to tackle such questions is spectroscopy: a technique that analyses the wavelengths of light coming directly from a planet's surface, or passing through its atmosphere. Each element or molecule produces a characteristic pattern of 'lines' — spikes of light emission or dips of absorption at known wavelengths — so observers can look at a distant object's spectrum to

read off what substances are present. "Without spectroscopy, you are to some extent guessing what you see," says Ian Crossfield, an astronomer at the University of Arizona in Tucson.

But spectroscopy has conventionally required a clear view of the object, which is generally not available for exoplanets. Most new worlds show up only as an infinitesimal dimming of a star as the otherwise invisible planet passes across its face; others are known only from the slight wobble of a star being tugged back and forth by the gravity of an unseen companion. Astronomers often say that trying to study such an object is like staring into a far-off searchlight (the star) and trying to see a firefly (the planet) hovering nearby.

In recent years, however, observers have begun to make headway. Some have extracted the spectra of light passing through the atmospheres of exoplanets as they cross the face of their parent stars — the equivalent of measuring the colour of the firefly's wings as it flits through the searchlight beam. Others have blocked the light of the parent star so that they can see exoplanets in distant orbits and record their spectra directly.

In the past two years, astronomers have begun to record spectra from a new generation of custom-built instruments such as the Gemini Planet Imager on the 8.1-metre Gemini South telescope at the summit of Cerro Pachon in Chile. Exoplanet spectroscopy will be a priority for several

spacecraft and ground-based telescopes that are now in development. And astronomers are waiting eagerly for NASA's James Webb Space Telescope (JWST), which will bring unprecedented light-gathering power and sensitivity to the task when it launches in 2018.

These are heady times for those hoping to get a deep understanding of new-found worlds, says Thayne Currie, an astronomer at Japan's Subaru Telescope on Mauna Kea, Hawaii. "We are on the cusp of a revolution."

TRANSIT SPECTROSCOPY

The first exoplanet in orbit around a Sun-like star was discovered in 1995, when astronomers Michel Mayor and Didier Queloz of the Geneva Observatory in Switzerland detected a regular, back-and-forth wobble in the movement of star 51 Pegasi. They concluded¹ that it was caused by the gravity of a planet at least 150 times the mass of Earth — roughly half the mass of Jupiter — orbiting the star every 4 days or so. Other discoveries followed as exoplanet fever took hold, and led telescope managers to make more observing time available for planet-hunting.

The list of finds soon sparked an idea for astronomer David Charbonneau of the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts. He reasoned that when a planet 'transits', or passes in front of a star, molecules in its atmosphere will absorb some of the starlight, and leave their spectroscopic fingerprints in it. Might it be possible to detect those fingerprints?

To find out, Charbonneau decided to look for sodium. "It's not particularly abundant," he says, "but sodium has very clear spectroscopic features" — excited molecules of it emit two very strong lines of light, which give sodium street lights their familiar yellow-orange colour. When the sodium is backlit, the light that floods through it has dark bands at the same points of the spectrum, and Charbonneau hoped that these would be comparatively easy to spot.

They were: in 2002, Charbonneau and his co-workers announced² that they had used the Hubble Space Telescope to detect a sodium signal from a Jupiter-sized exoplanet transiting HD 209458, a star about 47 parsecs (150 light years) from Earth. It was both the first detection and the first spectroscopic measurement of an exoplanet atmosphere. Within a few years, space-based transit observations were recording more complete spectra, and detecting gases such as carbon monoxide and water vapour.

Using this technique means looking for very tiny changes in a star's spectrum, says Charbonneau — maybe 1 part in 10,000. Hubble was and is observers' first choice of instrument: it does not have to contend with absorption of light by gases in Earth's atmosphere, so its spectra are very clean and easy to interpret. But competition for observing time is intense, so astronomers also use ground-based telescopes.

These do have to deal with atmospheric interference, but can overcome it by collecting more light than Hubble can. This allows them to detect fainter objects and to separate individual spectral features more clearly. That pays off because most exoplanets are in star systems that are moving relative to Earth. "So their wavelengths are Doppler-shifted," says Charbonneau, meaning that the radiation coming from them is stretched or squeezed by their movement, displacing the spectral lines slightly from the corresponding lines in Earth's atmosphere. Because the two sets of spectral lines no longer overlap, observers can know for sure how much of the signal comes from the exoplanet. Using this method, astronomers have been able to detect gases making up as little as 1 part in 100,000 of a planet's atmosphere.

An extension of the transit-spectroscopy technique has allowed astronomers to measure the light reflected from a planet's face. They do this after the planet moves across the face of its star, when it will be on the far side of its orbit, with its daylight side facing Earth (see 'Star shades'). Observers will not be able to see it as a separate object — but they will know that its spectrum is combined with that of the

star, says Nicolas Cowan, an astronomer at the McGill Space Institute in Montreal, Canada. Shortly afterwards, however, the planet will pass behind the star and be eclipsed — at which point, says Cowan, "you go from a planet and star to just a star. If you measure the difference in flux, you can tell how much light comes from the planet." The process is demanding, he says, but it can measure the infrared spectra of a Jupiter-sized planet in a close orbit even if it is less than 0.1% as bright as the star.

An even more ambitious application of this technique is to follow an exoplanet through a complete orbit. By subtracting the star-only spectrum obtained during the planet's eclipse, observers can get spectra of the planet's atmosphere as its silhouette changes from a thin crescent just after transit to a half-moon shape as it swings to the side, then a full-face view on the far side. This allows them to produce a comparatively fine-grained map of the atmosphere and how it changes over time. Cowan and his co-workers first reported³ using this technique in 2012, with infrared data from NASA's Spitzer Space Telescope. They showed that the exoplanet HD 189733b was hottest within about 10 degrees of its equator, as predicted. Since then, other researchers have used Hubble and Spitzer⁴ to map exoplanet atmospheres in more detail. And Cowan says that with the JWST, "it will be easy to make a 3D map of the atmosphere of a hot Jupiter."

Transit spectroscopy does have its limitations. Some exoplanets have nearly featureless spectra characteristic of clouds, which consist of droplets or fine dust particles that do not leave their imprint on the spectrum in the same way as isolated molecules⁵. The clouds are a big headache, says Charbonneau. "We don't have any direct measurement of what the clouds are made of. We just know they block the light." They aren't necessarily made of water vapour. Charbonneau points out that the cloud-shrouded super-Earth GJ 1214b, 12 parsecs from Earth, is so hot that its clouds could be made of zinc sulfide and potassium chloride. On still hotter worlds, the clouds could contain droplets of iron or rock.

Lisa Kaltenegger, director of the Carl Sagan Institute at Cornell University in Ithaca, New York, points to another limitation of the transit method. "When light hits a transiting planet, it isn't just absorbed," she says. "It also gets bent in the atmosphere", making it impossible for an observer on Earth to see. This bending, known as refraction, increases as the atmosphere becomes thicker. If alien astronomers were trying to get a spectroscopic reading of Earth, she says, refraction would prevent them

from probing any deeper than 10 kilometres from the surface⁶. But most of Earth's water is in the lowest 10 kilometres of its atmosphere, she says — so by analogy, "water is going to be one of the hardest things to find in an Earth-like exoplanet".

DIRECT IMAGING

An alternative approach to finding and studying exoplanets is trying to block out the starlight and image them directly, the equivalent of looking for the firefly by holding a hand in front of the searchlight. Early efforts to do this were futile: even the dimmest parent star is much brighter than an exoplanet. The secret of success is to seek brighter fireflies wandering well away from the searchlight — that is, young planets still glowing from the heat of formation, in orbits far from their stars. The first directly imaged exoplanets were announced by two groups simultaneously in 2008. The objects included 3 planets about 60 million years old orbiting the star HR 8799 (ref. 7), and a single planet more than 100 million years old orbiting Fomalhaut (ref. 8), a bright star some 8 parsecs from Earth.

To obtain the spectra of such objects, astronomers turned to adaptive optics, a technology that corrects for the twinkling of a star caused by turbulence in Earth's atmosphere and makes it much easier to spot any exoplanets in its vicinity. Also essential are discs inserted into the telescope's optical pathway to block light from the star, and sophisticated signal processors to digitally sharpen the images.

"Direct-imaging spectra are beautiful and tell you a lot about the

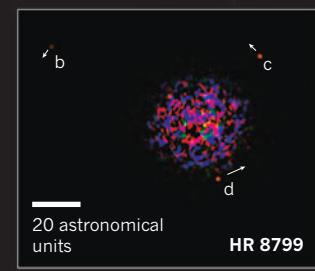
STAR SHADES

TRANSIT METHOD ▶

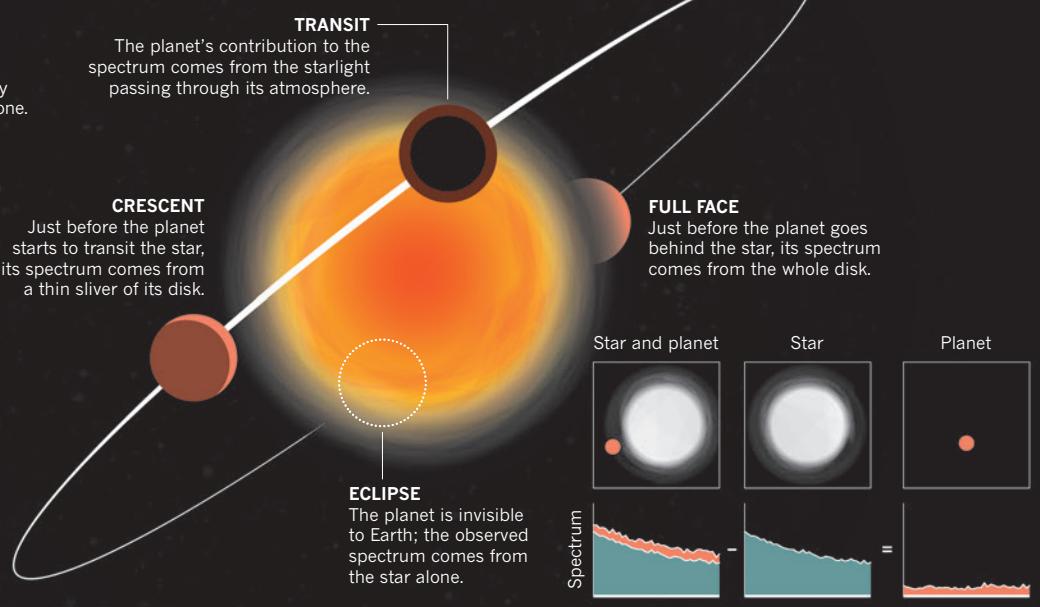
If a planet crosses the face of its star, astronomers can obtain its spectrum by subtracting the spectrum of the star alone.

DIRECT IMAGING

A disc inside the telescope can block enough of the star's light to allow imaging of planets in large orbits.



To analyse the light from an exoplanet, astronomers must first avoid having their instruments blinded by its much brighter star. There are two main ways of doing that.



SOURCE: SPECTRUM, NASA/JPL-CALTECH/R. HURT (SSC/CALTECH); DIRECT IMAGING, C. MAROIS ET AL./NRC CANADA

planets and how they formed," says Bruce Macintosh, an astronomer at Stanford University in California and a co-discoverer of the HR 8799 planets. In 2011, he and his colleagues reported⁹ the first detection of water vapour on one of those planets using a first-generation direct-imaging instrument that could observe only exoplanets with temperatures higher than 1,000 kelvin. Now, Macintosh is the principal investigator for the Gemini Planet Imager, which, along with the similar Spectro-Polarimetric High-contrast Exoplanet Research (SPHERE) imager at the European Southern Observatory's Very Large Telescope in Chile, is a second-generation instrument built to directly image and take spectra of exoplanets down to about 600 kelvin.

The Gemini instrument launched a multiyear search for Jupiter-like planets orbiting hot, young stars in November 2014. Early observations of 51 Eridani, a 20-million-year-old star about 30 parsecs away, spotted a Jupiter-like world 2.5 times farther from the star than Jupiter is from the Sun¹⁰. The spectrum showed that this exoplanet, dubbed 51 Eridani b, has an atmosphere containing more methane — a known component of Jupiter's atmosphere — than any other exoplanet. "The really exciting thing with 51 Eridani b and other new exoplanets," says Currie, "is that we see them when their spectra look a little more normal" and Jupiter-like than those of planets that are even younger and hotter, where methane is strangely absent. That could provide crucial insight into planet formation, the current theory of which is based mostly on data from the Solar System.

SPHERE has embarked on a similar survey, but started later, in February 2015, and has less to report. Thus far, says team member Anthony Boccaletti, an astronomer with the Paris Observatory, the most interesting discovery¹¹ is a group of five gas clumps moving at high velocity away from the young star AU Microscopii, which is known to be unusually prone to flares and other activity. "We don't really know what they are," he says.

STAR SURVEYS

Exoplanet spectroscopy has come a long way from its early days, when practitioners were struggling to extract extremely faint signals from noisy environments. The first results were often problematic. Now, Crossfield says, "for the most part what we are finding holds up and is repeatable".

A coming generation of instruments promises to reveal even more. NASA's Transiting Exoplanet Survey Satellite (TESS), scheduled to launch in August next year, will spend two years searching for exoplanets transiting more than 200,000 of the brightest stars in the solar neighbourhood. Exoplanets will also be targets for the JWST. With its

6.5-metre telescope and advanced instruments, Webb should see many more than the 2.4-metre Hubble. "TESS and Webb will own this space in five years," predicts Macintosh.

Two other planned — but not yet approved — space missions will use exoplanet spectroscopy. NASA's 2.4-metre Wide Field Infrared Survey Telescope, expected to launch in the mid-2020s, would spend most of its time on cosmological questions, but is expected to find and study about 2,600 exoplanets. Currie says that it should be able to image Jupiter-like planets orbiting nearby stars, although smaller, colder bodies similar to Pluto or the hypothetical 'Planet X' speculated to exist at the edge of the Solar System — or Earth, for that matter — will remain out of reach. "We would need a 10-metre-scale telescope in space to do other Earths," says Macintosh.

The second mission is ARIEL, the Atmospheric Remote-Sensing Infrared Exoplanet Large-survey, one of three candidates for a medium-class mission to be launched by the European Space Agency in 2026. The 1-metre telescope would be dedicated to transit spectroscopy and a survey of exoplanets at temperatures higher than 500 kelvin.

In about a decade, astronomers hope to see the completion of three super-giant telescopes: the 24.5-metre Giant Magellan Telescope at the Las Campanas Observatory in Chile, the Thirty-Meter Telescope planned for Mauna Kea, and the European Extremely Large Telescope on Cerro Armazones in Chile. All three will be equipped with adaptive optics systems, and it's a safe bet that they will be doing exoplanet spectroscopy to test models based on the data gleaned up to that point.

Those measurements could be astronomers' first realistic chance to find life in the wider Universe, says Charbonneau. "I'm so excited." ■

Jeff Hecht is a freelance writer in Auburndale, Massachusetts.

1. Mayor, M. & Queloz, D. *Nature* **378**, 355–359 (1995).
2. Charbonneau, D., Brown, T. M., Noyes, R. W. & Gilliland, R. L. *Astrophys. J.* **568**, 377–384 (2002).
3. Majeau, C., Agol, E. & Cowan, N. *Astrophys. J.* **747**, L20 (2012).
4. Wong, I. *et al.* Preprint at <http://arxiv.org/abs/1512.09342> (2015).
5. Sing, D. K. *et al.* *Nature* **529**, 59–62 (2016).
6. Bétrémy, Y. & Kaltenegger, L. *Astrophys. J.* **791**, 7 (2014).
7. Marois, C. *et al.* *Science* **322**, 1348–1352 (2008).
8. Kalas, P. *et al.* *Science* **322**, 1345–1348 (2008).
9. Barman, T. S., Macintosh, B., Konopacky, Q. M. & Marois, C. M. *Astrophys. J.* **733**, 65 (2011).
10. Macintosh, B. *et al.* *Science* **350**, 64–67 (2015).
11. Boccaletti, A. *et al.* *Nature* **526**, 230–232 (2015).

COMMENT

PHYSICS Two takes on our relationship with light, scientific and cultural **p.278**

ENGINEERING US roads are paved with questionable intentions **p.279**

ENVIRONMENT India soon to be overwhelmed by illegally imported e-waste **p.281**

OBITUARY Marvin L. Minsky, impish founder of AI, remembered **p.282**



ANDREW HOLBROOK/CORBIS



A shipbreaking yard in Chittagong, Bangladesh. Several workers in the country's scrapping industry are injured every week.

Three steps to a green shipping industry

It is time to crack down on the emissions and destructive development caused by vast container vessels that pollute the air and seas, write **Zheng Wan** and colleagues.

On 26 April 1956, US entrepreneur Malcom McLean watched a converted oil tanker leave Port Newark in New Jersey carrying 58 of his inventions: the modular shipping container. By 2015, the largest container ship in the world, with a deck the area of 3.5 soccer fields, could carry about 20,000 of the units.

Ever-bigger container ships carry 90% of global consumer goods such as clothes and food (non-bulk cargo)¹. The seaborne container trade has grown from 100 million

tonnes in 1980 to about 1.6 billion tonnes in 2014. Standardized 20-foot (6-metre) containers are moved using automated systems that connect seaports, airports and train stations². Bigger ships carry more containers, ideally consuming less oil and releasing fewer pollutants for each unit of goods carried.

Nonetheless, the human and environmental costs of shipping are vast. Low-grade marine fuel oil contains 3,500 times more sulfur than road diesel. Large ships pollute the air in hub ports, accounting for one-third

to half of airborne pollutants in Hong Kong, for example³. Particulates emitted from ships cause 60,000 cardiopulmonary and lung-cancer deaths each year worldwide⁴. Expanding harbours to take vast ships destroys coastal ecosystems. And scrapping fleets of obsolete smaller ships pollutes seas and soils, and damages workers' health, especially in the developing world⁵.

The industry is at a crossroads. The expected profits from larger ships are being undermined by excess capacity, slowing ▶

► trade and plunging transport prices. In 2015, container freight rates for the world's busiest shipping route — between Asia and northern Europe — dropped by nearly 60% in three weeks. A dozen shipping companies went bankrupt, including Denmark's Copenhill and China's Nantsing. Even the giant container-conveying Danish conglomerate Maersk announced that it would lay off 4,000 employees by 2017 and delayed or cancelled orders to build mega-ships.

Companies face a dilemma. If they buck the trend of scaling up, they risk being less competitive. Yet running mega-ships only part full wipes out the benefits of economies of scale. Ships use more fuel per container when half-loaded than for a full cargo.

The future is green shipping: efficient marine transport with minimal health and ecological damage⁶. Cleaner practices — especially on ship scrapping, emission control and port management — are needed. Achieving this will require heroic efforts by the industry and its engineers in collaboration with regulators, port authorities and communities. Environmental impacts should be considered in determining optimal routes and modes for delivery of goods.

POLLUTION PROBLEM

Shipping is the most energy-efficient way to move large volumes of cargo. Yet ships emit nitrogen oxides (NO_x), sulfur oxides (SO_x), carbon dioxide and particulate matter (PM)

into the atmosphere. Worldwide, from 2007 to 2012, shipping accounted⁷ for 15% of annual NO_x emissions from anthropogenic sources, 13% of SO_x and 3% of CO_2 . In Europe in 2013, ships contributed 18% of NO_x emissions, 18% of SO_x and 11% of particles less than 2.5 micrometres in size ($\text{PM}_{2.5}$). For road transport, the figures were 33%, 0% and 12%, respectively. Aviation, by contrast, accounted for only 6%, 1% and 1%, respectively, and rail just 1%, 0% and 0%.

Shipping policies must be applied worldwide to be effective. Shipping and aviation emissions are not addressed by global climate-change agreements, including the deal made in Paris last December. The International Maritime Organization (IMO), which regulates international shipping, is engaging — slowly. Releases into the oceans of oils, noxious liquids, harmful substances, sewage and garbage have been restricted since the 1980s by the International Convention for the Prevention of Pollution from Ships (MARPOL), following a spate of oil-tanker accidents. Air-pollution limits for shipping were adopted in 1997 but came into force only in 2005.

Energy efficiency is the IMO's present focus. Starting in 2013, its Energy Efficiency Design Index and Ship Energy Efficiency Management Plan aim to lower CO_2 emissions from shipping through tighter technical requirements on engines and equipment, maintenance regimes and voyage plans. No

absolute emissions-reduction targets were set. Unfortunately, long-term expansion in global trade and growing ship numbers mean that even if these measures are fully implemented, total shipping emissions are projected to quadruple from 1990 to 2050⁸.

The IMO has set up four 'emission-control areas' — the Baltic Sea, the North Sea, the US Caribbean and the coastal waters of Canada and the United States — where ships are required to minimize emissions mainly of SO_x and NO_x . These regions exclude the world's ten largest container ports, such as the Chinese ports of Shanghai, Shenzhen, Hong Kong and the South Korean port of Busan, which are all in Asia (see 'The dirty ten'). We estimate that these ten sites alone contribute 20% of port emissions worldwide.

A few developed countries, including the United States, the United Kingdom and Norway, limit the sulfur content of marine fuel in their national waters to within 1,000 parts per million (p.p.m.). Most developing countries, including India and China, permit dirtier fuels with 35,000 p.p.m. of sulfur. The European Union fuel standard for cars is 10 p.p.m.

Ship scrapping is heavily polluting. Asbestos, heavy metals and oils are toxic. Workers are exposed to hazardous fumes. The EU has laws requiring that ships registered in Europe be broken up only in licensed yards that meet strict guidelines. But it is easy to change a ship's registration and demolish it in a country with a more lax approach to labour and environmental protection.

India, Bangladesh, and Pakistan are popular for ship scrapping⁹. In Bangladesh for example, 40,000 mangroves — trees that stabilize many tropical coasts and are habitats and breeding grounds for many species — were chopped down in 2009 alone to accommodate shipbreaking yards. The pollution from scrapping there has caused an estimated 21 fish and crustacean species to become extinct. And reportedly, each week one worker dies and seven are injured in the scrap yards of Bangladesh.

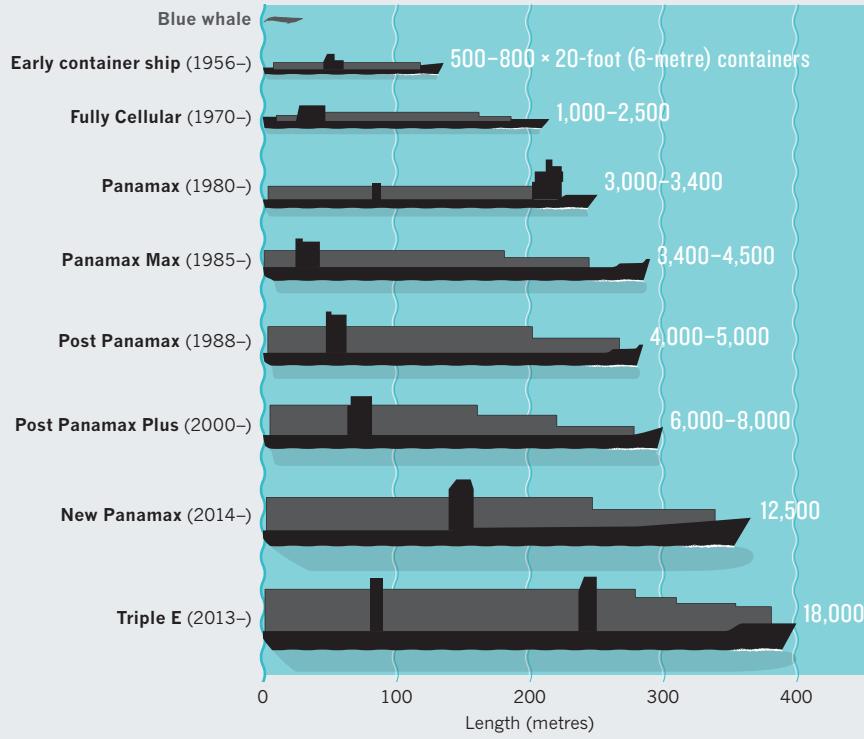
Congestion adds to pollution and disruption. Large volumes of cargo overwhelm ports, surrounding roads and waterways. Hasty expansion or construction of berths and canals to take more large ships can be environmentally disastrous. Where the water in existing harbours is too shallow, port authorities may reclaim land from the sea or build artificial islands in deeper waters.

Coastal changes destroy ecosystems. Over the past three decades, about 75% of mangroves have disappeared from Shenzhen, following port expansion and land reclamation. Plans for the Porto Sul port in Brazil — slated to open in 2019 — identified 36 potential environmental impacts, including driving away dolphins and whales and killing seabirds.

Traditional shipping routes cannot keep

SUPERSIZE SHIPS

Each generation of container ship is getting bigger as economies of scale are expected to bring down transportation costs. But the largest rarely carry a full cargo, and pollute hub ports.

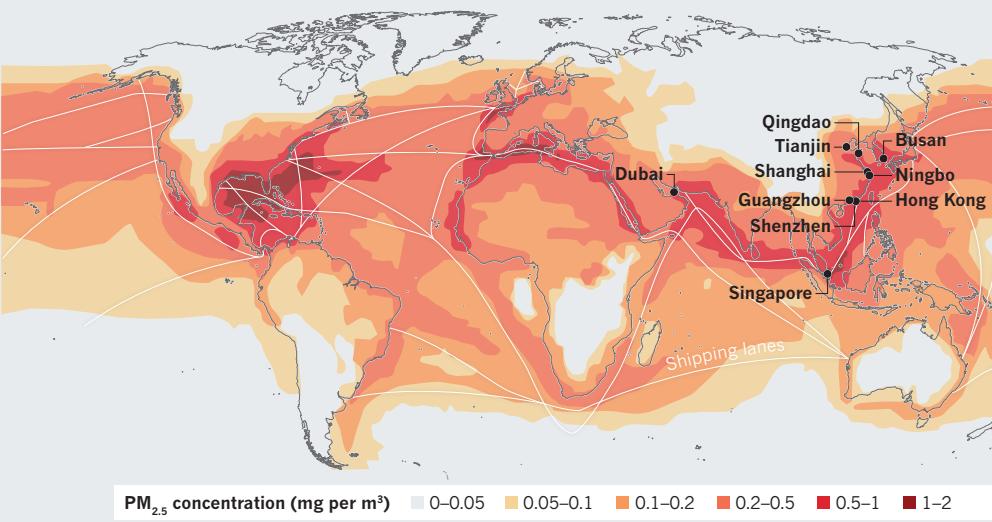


SOURCE: J.-P. RODRIGUE, THE GEOGRAPHY OF TRANSPORT SYSTEMS, 3RD EDN (ROUTLEDGE, 2013) (SHIP-CAPACITY DATA)

THE DIRTY TEN

Particulate matter less than 2.5 micrometres ($PM_{2.5}$) emitted from dirty marine fuel oil causes poor air quality along shipping lanes. Emissions-control zones omit the ten largest container ports, which contribute an estimated 20% of worldwide port emissions of nitrogen oxides and sulfur oxides.

COMMENT



SOURCE: REF. 4 ($PM_{2.5}$ CONCENTRATIONS); J. COMMERCE
TOP 50 WORLD CONTAINER PORTS 2014 (PORTS)

up. The Panama Canal, which connects the Pacific and Atlantic oceans, can currently handle vessels carrying only up to about 5,000 standard containers. A project to expand it to accept ships with 13,000 containers (the 'New Panamax' class) should be completed by May. But the largest megaships, such as Maersk's E-class and Triple E-class (with capacities between 14,000 and 18,000 containers), will still be unable to cross (see 'Supersize ships'). In the meantime, heavy traffic at Panama, complicated navigation and constant maintenance have led to a ten-day delay in voyage times.

To take advantage of the business opportunity, construction is scheduled to start this year on a 280-kilometre-long canal through Nicaragua. This US\$50-billion project, funded by a billionaire-owned Hong Kong company, could destroy almost 400,000 hectares of tropical forests and wetlands, home to threatened and endangered wildlife and indigenous communities¹⁰.

Public concern about the pollution and health impacts of shipping remains muted because the industry is a backbone of the global economy, and its activities happen far from where most people live and often beyond the jurisdiction of local regulators. We cannot rely only on new ship designs and engine innovation to minimize the ecological footprint of shipping: today's ships might be in use for another 20 years or more. Several issues must be addressed together to make the industry greener.

GREEN SHIPPING

Implementing the following recommendations could save thousands of lives each year, ensure cleaner coastal air and reduce ecological damage from shipping.

Clean up ship scrapping. The IMO adopted the Hong Kong International Convention for the Safe and Environmentally Sound Recycling of Ships in 2009, but only Norway,

Congo and France have acceded as of February 2016. The IMO's priority should be to ensure that the principal scrappers — India, Bangladesh and Pakistan — adhere to these guidelines. The first step is to set up local offices in these countries to collect and analyse monitoring data independently and to propose improvements to local governments. International loan or aid programmes to these countries, sponsored by the World Bank or the Asian Development Bank, for example, should demand clean ship-scraping practices as an incentive. To discourage transfer of scrapping elsewhere, a watch list of poorly performing countries needs to be updated by IMO regularly until an international convention enters into force.

Control emissions. Stricter IMO emissions regulations are needed, including a cleaner worldwide standard for sulfur released by combustion of marine fuel. A 97% cut in SO_x can be achieved by reducing the sulfur content from 35,000 p.p.m. to 1,000 p.p.m. fuel oil. Today's low oil price provides a great opportunity for this transition to happen. The current cost of 1,000-p.p.m.-grade fuel oil (around US\$300 per tonne in Singapore, for example) is less than half of that of the cheapest dirty fuel four years ago.

Marine fuel is a sideline for oil refineries — only 2–4% of the total fuel market. Stricter emissions standards will stimulate demand for high-quality fuel. Incentive programmes (tax rebate and subsidies for producers) will be needed to ensure a reasonable profit margin to recover the initial high investment in developing countries, where there is little current capacity. Government interventions will be needed in countries with state-run oil companies, such as in China and India.

An alternative is to install scrubbers for exhaust-gas cleaning on ships. Scrubber units blend the exhaust gas with water or caustic soda to remove up to 99% of SO_x and 98%

of particulate matter from high-sulfur fuel. At the moment, scrubbers are expensive, costing \$2 million for one ship. But China, for instance, could equip its entire container fleet in one year by funding a 50% subsidy for scrubbers. The total cost? Just 0.5% of the \$150 billion per year it has spent since 2013 to fight pollution. Shipping companies could recoup the other 50% in one year from fuel savings. With a stricter emissions standard, the demand for scrubbers would go up, and the costs down, as production scales.

Improve port management. Port authorities should review the environmental impact of their previous construction and disclose information on their future development plans to demonstrate responsible management of public assets. They should coordinate with transport-planning bureaus to seek the most economical and environmentally friendly strategy to dispatch goods; the optimal capacities of its terminals; and how to assist ships to load and unload quickly. Making port-business statistics and the results of environmental-impact studies accessible will allow the research community to be involved in the decision-making process. Environmental non-governmental organizations should campaign to increase public awareness of port development.

After decades of loose oversight, it is time for shipping to get a whole lot greener. ■

Zheng Wan is associate professor, **Mo Zhu** is assistant professor and **Shun Chen** is associate professor at the College of Transport and Communications, Shanghai Maritime University, Shanghai, China. **Daniel Sperling** is distinguished professor of civil engineering and environmental science and policy, and founding director of the Institute of Transportation Studies, at the University of California, Davis, USA. e-mail: mrwan@ucdavis.edu

1. Singh, S. P., Saha, K., Singh, J. & Sandhu, A. P. S. *Packaging Tech. Sci.* **25**, 149–160 (2012).
2. Levinson, M. *The Box: How the Shipping Container Made the World Smaller and the World Economy Bigger* (Princeton Univ. Press, 2010).
3. Hong Kong Environmental Protection Department. *2013 Hong Kong Emission Inventory Report* (Hong Kong Government, 2015); available at <http://go.nature.com/qijuuu>
4. Corbett, J. J. et al. *Environ. Sci. Tech.* **41**, 8512–8518 (2007).
5. Demaria, F. in *Nature, Economy and Society: Understanding the Linkages* (eds Ghosh, N., Mukhopadhyay, P., Shah, A. & Panda, M.) 273–304 (Springer, 2016).
6. Polemis, S. M. *Balancing Environmental Aspirations with Economic Realities* (International Chamber of Shipping, 2012).
7. Smith, T. W. P. et al. *Third IMO Greenhouse Gas Study 2014* (International Maritime Organization, 2014).
8. Anderson, K. & Bows, A. *Carbon Mgmt* **3**, 615–628 (2012).
9. Hossain, K. A. J. *Environ. Anal. Toxicol.* **5**, 312 (2015).
10. Huete-Pérez, J. A. & Meyer, A. *Nature* **506**, 287–289 (2014).

Radiant realms

Philip Ball enjoys two explorations of light, spanning wonders from Newton's spectrum to the aurora borealis.



The northern lights over Norway — produced by the solar wind and Earth's magnetic field.

Last year's International Year of Light is now history, but light remains high on the scientific agenda. From photosynthesis to quantum optics and light-emitting organic compounds, there is no getting away from it, even if you wanted to. Bruce Watson's *Light* and Melanie Windridge's *Aurora* — one encyclopaedic, the other more focused — are as eloquent about the aesthetic pleasures of light as they are about the

scientific issues that it presents. The temptation to call them lucid and illuminating only highlights (there I go again) the linguistic currency of this fundamental phenomenon.

In *Light*, Watson offers a wide-ranging tour of the cultural, literary and scientific response to his subject. Our associations with light seem almost unrelentingly favourable. Light is divine, dark demonic. "God Appears & God is Light/To those poor Souls who dwell in

Night", wrote William Blake in 1803, in 'Auguries of Innocence'. What could seem more natural than this dualism? We have sought for centuries to understand and use light, and never stopped worshipping it. Watson seems to acknowledge that by bookending his study with contemporary sunrise ceremonies at the prehistoric stone monuments of Stonehenge in England and Newgrange in Ireland.

The stops along the way will surprise no one who has a passing acquaintance with the traditions and science of light. The dazzling list includes creation myths, Aristotle's meditations on vision and ether, the light metaphysics of medieval Neoplatonism that identified God with radiance and informed Gothic architecture, the tonal contrast of light and dark (chiaroscuro) developed by Leonardo da Vinci, the work of early-modern scientists from Johannes Kepler to Isaac Newton, polymath Thomas Young's wave theory of 1800, painter J. M. W. Turner's last words ("The Sun is God"), Thomas Edison's light bulb, Albert Einstein's quanta and the laser. Even when covering familiar ground, Watson's touch is lyrical and deft.

Romantic poet John Keats famously declared in his 1814 epic poem *Lamia* that the progression from uncomprehending wonder to scientific understanding had unwoven the rainbow. He underestimated light's enduring appeal. Light might now be just a little slice of a continuous electromagnetic spectrum; trapped in optical fibres and sent hither and thither at our bidding; chopped into quanta and parted into electrical and magnetic oscillations. But it has not lost its primeval power to enchant. Modern artists make astonishing light sculptures, from James Turrell's subliminal illusions to Carlos Cruz-Diez's disorienting chromatically saturated light-rooms. Coloured light is still used to summon contemplation, as seen in the stained glass of Henri Matisse's mid-twentieth-century Chapel of the Rosary in Vence, France. And the crowds that Watson joined at Neolithic monuments are not hippie throwbacks. They are simply people eager to experience the awe of a primal event — the appearance of the Sun's first rays.

In any case, it is foolhardy to assert that light is scientifically understood. It is more like a beam guiding us to deeper mysteries. Newton's experiment with prisms in a "darkend chamber" to reveal the spectrum of sunlight is often presented as a turning point, celebrated by Alexander Pope in his 'Epitaph on Sir Isaac Newton': "Nature and Nature's Laws lay hid in Night/God said, "Let Newton be!" and all was light." But physicists including Young, James Clerk Maxwell and Einstein all radically altered Newton's picture — a picture that is still shifting, with the use of quantum optics to explore the fundamentals of quantum theory and reveal the strange connections that may persist between photons.

Light: A Radiant History from Creation to the Quantum Age
BRUCE WATSON
Bloomsbury: 2016.

Aurora: In Search of the Northern Lights
MELANIE WINDRIDGE
William Collins: 2016

Watson's book is an eye-catching display, reflecting and refracting like a gemstone. One moment we are among late-nineteenth-century Parisians watching the cinematograph of the Lumière brothers (was there ever a better example of nominative determinism?). The next we are examining Rembrandt's 1642 painting *The Night Watch* — which depicts a militia emerging into the light — or commiserating with modern astronomers about light pollution. This is all tremendous fun, but never quite focuses into a coherent image. Perhaps light has too many facets to offer unifying themes. However, like other mythologically charged substances (such as water and gold), it does not easily shed its associations under scientific scrutiny. That may be why transformation optics has become a 'technology of invisibility' and quantum optics claims to promise teleportation. The electromagnetic ether was once considered a possible bridge between the physical and spiritual worlds; in a way, light still holds that allure.

Nowhere is that more apparent than in the phenomenon of the aurora borealis, or northern lights, produced when subatomic particles in the solar wind collide with molecules in the atmosphere at Earth's magnetic pole, stimulating emissions of light. Plasma physicist Melanie Windridge embraces the aurora's magical quality even as she attempts to unravel the science. She has camped at temperatures approaching -40°C in Svalbard,



Sunlight streaming through windows at New York City's Grand Central Terminus.

Norway's Arctic archipelago, to experience the visceral thrill of seeing these shimmering curtains. It is clear in this captivating book that her technical understanding has not dimmed her delight.

Windridge's account explains what we do and do not yet know about the auroras, and how this understanding has evolved. The journey takes in the restless, violent activity of the Sun's outer layers, the physics of interaction between the plasma of the solar wind and Earth's magnetic field, the atomic physics that generates the various colours of the aurora (oxygen produces the predominant green and red, nitrogen the rarer blue and purple) and the potential of auroral monitoring (for

instance, of changes in shape) to provide advance warning of disruption to telecommunications caused by stormy "space weather".

What emerges is the tremendous difficulty of formulating intuitive and predictive models of a turbulent and chaotic process. The auroral morphologies are described even now in qualitative terms — arcs, veils, rays and so forth — and a full theory is still lacking. Windridge makes a persuasive case that the quest to find one is both pragmatic and inspired by the visual majesty of nature's light. ■

Philip Ball is a writer based in London. His latest book is *Invisible*.
e-mail: p.ball@btinternet.com

ENGINEERING

Broken bridges and highways from hell

Kyle Shelton applauds a study that probes the parlous state of US infrastructure.

In *The Road Taken*, engineer and historian Henry Petroski surveys the state of US bridges, roads and tunnels — the legacy of two centuries of technological development — and finds them crumbling. Echoing Robert Frost's poem 'The Road Not Taken', Petroski reflects on both physical highways and the choices that contributed to their current state. As Frost wrote, "Two roads diverged in a wood, and I—I took the one less traveled by,/And that has made all the difference." Petroski compellingly shows that only by closely investigating the figurative roads down which the country has travelled,

and the decisions made at their forks, can the past be used to shape the future.

Some of those paths, he reveals, have been revolutionary — from the advent of asphalt for road construction in 1870 to the quest for autonomous cars. Some have thrown up obstacles that affect future choices: expensive projects, huge maintenance demands,

© NATURE.COM
For more on science
in culture see:
[nature.com/
booksandarts](http://nature.com/booksandarts)

short-sighted political decisions and lacklustre design. Petroski shows how the choices of policymakers and engineers contributed

to soaring successes such as the Golden Gate Bridge in San Francisco, California, and tragic failures including the 2007 collapse of the I-35W bridge in Minneapolis, Minnesota, which killed 13 and injured 145.

A typical history of US transport infrastructure would begin with the travel revolution of the early nineteenth century. This spawned the steamship, New York's Erie Canal and the first federally funded highway, the National Road. Then followed railways, bicycles and the early-twentieth-century boom in automobiles and road construction.

Petroski touches on much of this, as well as

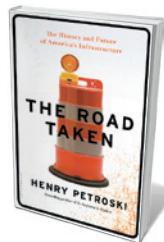
infrastructure below ground. In an opening vignette about his Brooklyn childhood, he highlights the contradictory nature of all that cabling and piping — simultaneously invisible and utterly necessary, the ever-present context of our daily lives, carrying electricity, waste, water and gas to and from our homes. But his focus is the structures and elements of the ever-evolving road.

He delves into the progression of road surfaces from carefully stacked stone to tarmac, concrete and asphalt, in the unending quest to make smooth, durable surfaces. He shares the history of modal conflict — between horse-powered vehicles, pedestrians and automobiles — by looking at how safety features and traffic signals developed. Finally, he discusses materials failure not only through how steel and concrete deteriorate, but by connecting that decay to financial and political choices that undervalued risk. There are more than 6 million kilometres of public roads in the United States, and its citizens travel nearly 5 trillion kilometres a year on them. But as the American Society of Civil Engineers reported in its 2013 'report card' on US infrastructure, 66,749 US bridges (1 in 9) are structurally deficient, and 32% of US roads are in "poor or mediocre condition".

Petroski shows how past practices ripple forwards. Details paint a rich, sometimes pungent, history: for example, woodblock roads (first used widely in the mid-nineteenth century) fell out of favour because they absorbed horse urine. On the country's notorious scourge of potholes, Petroski describes how the combined forces of heavy traffic, water and freezing temperatures conspire to create spreading cracks and distortions in asphalt. He concludes by casting forwards to future roads. Here, he opines, self-driving cars will create safer roads by calculating hazards faster than human drivers can. Asphalt could be revolutionized by mixing in fibres of steel wool; when these are heated by induction coils on a special vehicle, the asphalt will melt and reform, becoming self-mending.

Petroski's goal is to ask how, given the importance of the car to the US economy and mobility, federal and state governments have allowed the country's infrastructure to reach crisis point. But he goes beyond hand wringing. With an engineer's technical knowledge and a historian's eye, he offers a nuanced argument about the political, financial and engineering calculus that contributes to failures.

Acknowledging that infrastructure is part of the political domain,



The Road Taken:
The History
and Future
of America's
Infrastructure
HENRY PETROSKI
Bloomsbury: 2016



The heavily used I-35W bridge in Minneapolis, Minnesota, collapsed in 2007.

Petroski urges political engagement. He cites the controversial rebuilding of the eastern span of the Bay Bridge in Oakland, California, which opened in 2013. This cost a whopping US\$6.4 billion, took 12 years to build, was beset by construction problems and was arguably over-designed. He also discusses the disputes over the Department of Transportation's Highway Trust Fund, intended as the primary support for federal investment in US road systems, but due to run out of money this June. Politicians, he shows, too often choose expedience over sound solutions.

Despite that focus, Petroski does leave out many of today's crucial political debates. There is almost no discussion of the local decisions that shape US roads, such as where or how state and city officials decide to apply federal and state funds, or how municipalities shore up their road funds through different financing mechanisms. Nor does Petroski delve into the vexed issue of the displacements caused by interstate construction since the 1950s, which at its peak forced between 60,000 and 100,000 people to leave their homes every year. Finally, despite framing the book as a look at choices made

and not made, he does not comment on the possibility that roads and their continued construction are creating further problems.

As politicians and citizens consider how to fix existing highways, byways and bridges, they must also consider whether remaining so tethered to the system is a necessity. Might it not be time to ask how major investments in mass transit could complement, if not start to replace, the US reliance on cars? Reflective decision-making is, after all, at the core of what this informative book recommends. ■

Kyle Shelton is programme manager and a fellow at the Kinder Institute for Urban Research at Rice University in Houston, Texas. e-mail: kyle.k.shelton@rice.edu

CLARIFICATION

Surfaces, the exhibition reviewed in 'Medical modernist' (Nature 530, 30–31; 2016), was originally a joint production of the Museum of Concrete Art and the German Museum of Medical History, both in Ingolstadt.

Correspondence

Technology alone won't save climate

A dragon was buried at the Paris climate meeting (COP21): 'climate sceptics' disappeared. Now we face a second, equally formidable dragon: unreasonable optimism about 'new' energy technologies. This optimism supports economic-growth models driven by innovation, but depends on an unimaginable scale and rate of deployment.

Defeating the second dragon requires that we reconsider our habits of energy usage. Thirty years of engine-efficiency gains have been eclipsed by our preferences for ever-larger cars that are often 20 times heavier than the passengers — but these are habits, not needs.

We could continue to live well in rich economies with, say, one-quarter of the energy. For instance, we could run the boiler for one-quarter of the time and quarter our movement of mass — the total of all vehicles, freight and people, measured in tonne-kilometres. We could also make buildings and goods with half the material (without risking safety) and keep them for twice as long.

'Success' today is largely associated with derivative measures of increasing gross domestic product, profitability, speed or salary. Yet our value systems are based on integral measures of quality and stock: reputation, heritage, journeys and relationships. We need to expand the dialogue of climate mitigation to reflect these values. Challenging our habits of energy use should be the first priority of climate policy.

Julian Allwood *University of Cambridge, UK.*
jma42@cam.ac.uk

Formalize recycling of electronic waste

India urgently needs a formal recycling policy for its mountain of electronic waste. Boosted by illegally imported discards from

the West, this waste is expected to reach a total of around 30 million tonnes by 2020.

Western electronic waste comes largely from countries' weak legislation on its handling and management (G. Agoramoorthy and C. Chakraborty *Nature* **485**, 309; 2012). Although people in India informally recycle an estimated 95% of electronic waste for profit, the practice could soon be overwhelmed.

India's government proposed draft regulations for this waste in June 2015, to be formalized after a public consultation. These are already proving effective, but there is still a pressing need for national policy to alleviate damage to the environment. This would create employment and commercial opportunities, address health and safety concerns, and forge a path towards sustainability.

Devika Kannan, Kannan Govindan *University of Southern Denmark, Odense, Denmark.*
Madan Shankar Anna *University, Chennai, India.*
kgov@iti.sdu.dk

International accord on open data

The accord *Open Data in a Big Data World* has been produced by representative bodies of global science collaborating as Science International (see go.nature.com/tpq3tu). It sets out the principles for maximizing benefit from the digital data revolution in shaping the future conduct of science.

Openness is the bedrock for benefit. Whole science systems, not merely the habits of researchers, need to adapt. It will be necessary for public funders of research to fund open-data management, for publishers to ensure that open data are deposited concurrently with the publication of derived scientific claims, for disciplinary societies to debate how their disciplines should adapt, and for universities to create incentives and support for open-data processes.

The accord recognizes

potential pathologies: that the data deluge could overwhelm the open scrutiny of scientific claims, and that a countervailing trend towards privatization of knowledge could be at odds with the ethos of scientific inquiry and our need to use ideas freely.

It is crucial that standards of reproducibility are re-established for a data-rich age, and that the global scientific community commits to "intelligently open" science (see go.nature.com/dvgdfo). Digital technologies also provide a route to open science and open knowledge, where all sectors of society are involved in the co-design and co-production of actionable knowledge.

Geoffrey Boulton *ICSU CODATA; and University of Edinburgh, UK.*
g.boulton@ed.ac.uk

Control wildlife pathogens too

Policies to control diseases caused by invasive alien species should be extended to cover endangered wild species, ecosystems and their services — not just humans, livestock and cultivated plants.

Of the 100 invasive alien species listed by the International Union for Conservation of Nature as the 'world's worst', one-quarter have environmental impacts that are linked to diseases in wildlife (M. J. Hatcher *et al. Front. Ecol. Environ.* **10**, 186–194; 2012). Identifying and managing this threat calls for coordinated interdisciplinary expertise.

Priorities are to collect baseline information on the distribution and population dynamics of pathogens, hosts and vectors; to determine the relative importance of invasion pathways; and to develop methods for predicting host shifts, pathogen-host dynamics and the evolution of alien pathogens (see also go.nature.com/ux4wpp).

This integrated strategy is geared towards the goals set by the Convention on Biological Diversity for managing invasives.

Helen Roy* *NERC Centre*

for Ecology and Hydrology, Wallingford, UK.
hele@ceh.ac.uk

*On behalf of 4 correspondents (see go.nature.com/upyjwi for full list).

Weapons plutonium riskier above ground

Cameron Tracy and colleagues argue that the US Department of Energy (DOE) should conduct a new safety assessment of its nuclear Waste Isolation Pilot Plant (WIPP) before loading it with 34 tonnes of plutonium from dismantled nuclear weapons (*Nature* **529**, 149–151; 2016). We contend that the long-term risk of disposal in WIPP should be balanced against the benefits.

The reanalysis that the authors propose would take several years. During this time, Congress could well abandon the disposal programme. Leaving the plutonium above ground indefinitely would pose a much greater environmental threat than disposing of it in WIPP.

Disposal in WIPP would also offer a cheaper, simpler and more secure alternative (see go.nature.com/h81nb5) to the unaffordable plan to convert the plutonium to mixed oxide with uranium and burn it in nuclear-power plants.

There is a way to keep the disposal programme moving. The DOE proposes to put 6 tonnes of excess weapons-usable plutonium in WIPP from its Savannah River site. This will not markedly affect the WIPP inventory. Packaging that plutonium for disposal will take about 6 years (see go.nature.com/qhl8uf). Meanwhile, the DOE could redo the WIPP safety analysis and evaluate other disposal options for the 34 tonnes of plutonium (see, for example, go.nature.com/2cik4o), which will remain in bunkers at Savannah River and in Texas until a solution is found.

Edwin Lyman *Union of Concerned Scientists, Washington DC, USA.*
Frank von Hippel *Princeton University, New Jersey, USA.*
fvhippel@princeton.edu

Marvin L. Minsky

(1927–2016)

A founding father of artificial intelligence.

Marvin Lee Minsky had no patience for those who doubted that computers could be intelligent at a human level or beyond. In the late 1950s, building on the work of Alan Turing, along with computer scientists John McCarthy, Herbert Simon and Allen Newell, Minsky started the work that led everyone to think of this group as the founders of the field of artificial intelligence (AI). Were it not for their determined advocacy, AI might have founders.

Minsky, who died on 24 January, was born in New York City in 1927. After serving in the US Navy in the Second World War, he earned a degree in mathematics in 1950 from Harvard University in Cambridge, Massachusetts, where he impressed the mathematician Andrew Gleason by proving fixed-point theorems in topology. During his doctorate on learning machines, at Princeton University in New Jersey, he built one out of vacuum tubes and motors.

When Minsky finished his PhD, the eminent mathematicians John von Neumann, Norbert Wiener and Claude Shannon all recommended him for appointment as a junior fellow at Harvard. During this time, he became curious about how the brain works, but was frustrated by the limitations of conventional microscopy, which could not provide clear images of thick, light-scattering neural tissue. This led to his invention of the confocal scanning microscope, which uses lenses to focus light on successively small volumes.

In the late 1950s, with McCarthy, he founded a group that became the Artificial Intelligence Laboratory at the Massachusetts Institute of Technology (MIT) in Cambridge. In 1961, Minsky published his famous paper 'Steps toward artificial intelligence' (M. Minsky *Proc. IRE* **49**, 8–30; 1961), a call to arms for a generation of researchers. Scientists flocked to Minsky's laboratory to take on the challenge of understanding intelligence and endowing computers with it. They benefited from Minsky's wisdom and enjoyed his insights, lightning-fast analyses and impish repartee.

His students felt part of a scientific revolution. They helped Minsky to develop high-level theories about how programs could recognize structures made of toy blocks, answer questions about stories written for children, learn something definite from individual examples and exhibit common sense.



His laboratory was an egalitarian utopia. He didn't notice looks, gender, age or status. He cared only about ideas and ability. Minsky and his wife, Gloria, often welcomed students into their home, where several pianos stood as a reminder that Minsky was a musical prodigy, able to improvise fugues.

Minsky's attention span was short. Whenever I explained an idea to him, he would leap ahead of me, having worked the whole thing out after a few sentences. Once, I suggested that if we ever developed really intelligent machines, we should do a lot of simulation before we let them loose in our world to be sure they weren't dangerous. "And we're the simulation?" he said, guessing my punchline. "It isn't going very well, is it?"

His laboratory built pioneering robots as well as revolutionary programs. Minsky invented a robot arm with 14 degrees of freedom. He argued that space exploration and nuclear-material processing would be simpler with manipulators driven locally by computers or remotely by human operators. He foresaw that microsurgery would be done by surgeons by using telepresence systems.

In the late 1960s, Minsky and MIT mathematician Seymour Papert worked on the mathematics of perceptrons — simple neural networks — showing what they could and could not do, which raised the sophistication of research on neurally inspired mechanisms to a higher level. Minsky and Papert collaborated into the 1970s and early 1980s, developing theories of intelligence and radical approaches to early

education that centred on teaching children to program using the Logo language.

In the mid 1970s, Minsky introduced 'frames', a way of describing entities and situations using a template-like representation. A frame describing a birthday party, for example, would have a slot for the person celebrated, that person's age and a list of the gifts presented, along with slots for time and place inherited from a 'celebration' frame. He also developed the idea of knowledge lines (K-lines) to address questions about how information is represented, stored, retrieved and used in the brain. He argued that K-lines help us to solve problems by putting us back into mental states that resemble those we were in when we previously thought about similar problems.

In 1985, he brought these and many other ideas together in a book, *The Society of Mind* (Simon & Schuster). He wrote that intelligence emerges from the cooperative behaviour of multiple agents, none of which is intelligent. Then, in 2006, Minsky published *The Emotion Machine* (Simon & Schuster), a book about intelligence, creativity, emotion, consciousness and common sense. Multiplicity is a dominant theme. He noted, for example, that concepts such as intelligence are 'suitcase words', into which one can stuff multiple meanings. He wrote that our resourceful intelligence arises from multiple ways of thinking on multiple levels, and from multiple ways of representing knowledge.

In recent years, Minsky found it ironic that the doubters of the possibility of AI have been replaced by worriers about its consequences. He didn't see a technical advance that would justify the change in attitude, attributing recent successes in AI to faster computers. He thought that not much real progress had been made in the field for several decades, but he had, nevertheless, no doubt that our species' greatest legacy will be the intelligent computers that we create.

Minsky's talks, papers and books are like diamond mines. The riches will take decades to cut and polish, inspiring researchers for decades to come. ■

Patrick Henry Winston is professor of artificial intelligence at the Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. He was a graduate student of Marvin Minsky's in the 1960s, and thereafter an admiring friend and colleague. e-mail: phw@mit.edu

QUANTUM PHYSICS

Photons paired with phonons

The force exerted by light on an object has been used to pair photons with quantum units of mechanical vibration. This paves the way for mechanical oscillators to act as interfaces between photons and other quantum systems. **SEE LETTER P.313**

MILES BLENCOWE

Quantum entanglement¹ is a bizarre state in which it is meaningless to describe the properties of individual objects in a collection; only the properties of the collection as a whole may be described. On page 313 of this issue, Riedinger *et al.*² report the quantum pairing of light and vibrations of microscopic mechanical oscillators comprising more than 10^{12} atoms — large for a quantum object. This is a big step towards the goal of using light to achieve the quantum entanglement of the vibrational motion of two widely separated mechanical oscillators, aiding the development of quantum-information processing systems that have practical applications.

Riedinger and colleagues exploit the fact that light shining on an object exerts a force³. If the object is a freely suspended wire that is clamped at both ends, such as a silicon nanobeam, then an incident light pulse will set it vibrating, like tapping a bell with a hammer. The silicon nanobeam used in the authors' experiments was about 15 micrometres long, 500 nm wide and 250 nm thick, and was engineered so that a fraction of the incident light from a near-infrared laser source could be trapped in a segment of the nanobeam.

This segment functions like an optical cavity (a system used to trap light at certain frequencies known as resonances), in which the cavity length is comparable to the wavelengths of both the light and of a particular vibrational mode of the nanobeam. The force exerted by the light is considerably enhanced⁴ by co-locating the optical cavity and the vibrating regions, rather than letting the light 'tap' the nanobeam from the outside. The mechanical-vibration mode is driven from within by the trapped light, and manifests as a 'breathing' mode: rapidly alternating expansions and contractions of the beam's width at about 5.3 gigahertz.

At the very low light intensities used in Riedinger and co-workers' experiments, a quantum description of the system is necessary in which the light consists of photons and the vibrational energy of the nanobeam comes only in discrete lumps called phonons. Describing the action of the light force at the quantum level, an incident photon can emit and hence create a phonon only if the

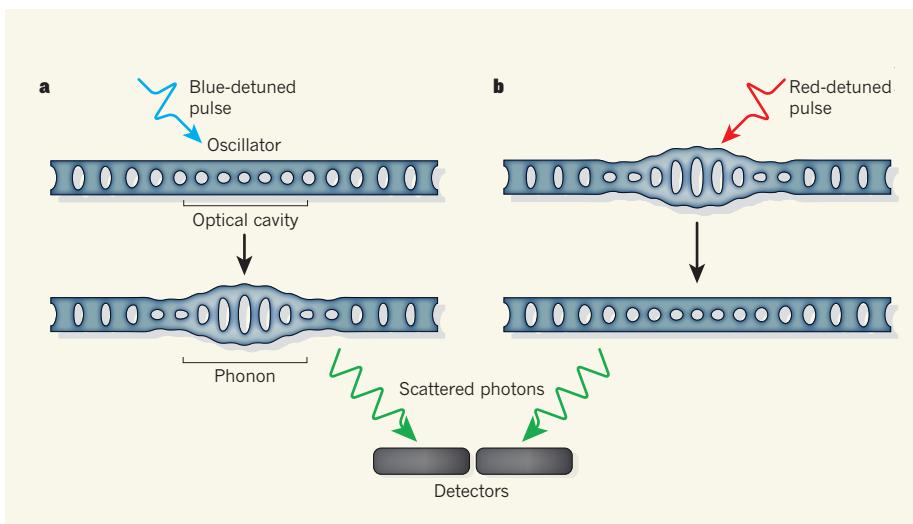


Figure 1 | A photon-phonon interface. **a**, Riedinger *et al.*² shone pairs of light pulses on a microscopic mechanical oscillator (a silicon nanobeam) that also incorporated an optical cavity — a section with different-sized holes that traps standing waves of light at a resonant frequency. The first light pulses in each pair were blue-detuned (their energy is slightly higher than the resonance frequency of the optical cavity) and could induce the formation of a single phonon (a quantum unit of vibration) in the oscillator and produce a scattered photon at the resonance frequency. **b**, The second pulses were red-detuned (with slightly lower energy than the resonance frequency) and could absorb the single phonon from the nanobeam, again generating a scattered photon at the resonance frequency. By measuring the joint probabilities of scattered photons produced from the two types of light pulse, the authors established that the number of photons in the cavity correlates with the number of phonons.

energy lost by the resulting scattered photon corresponds to the resonance frequency of the cavity³ (Fig. 1a). Similarly, an incident photon can absorb and hence annihilate a single phonon only if the energy of the resulting scattered photon corresponds to the resonance frequency of the cavity (Fig. 1b). Higher-energy incident photons that can emit phonons are said to be blue-detuned with respect to the cavity's resonance frequency, whereas lower-energy incident photons that can absorb phonons are red-detuned.

Riedinger *et al.* cooled the silicon nanobeam to a few hundredths of a kelvin, and verified that the vibrational breathing mode was in its quantum ground state, to a good approximation — the breathing mode did not contain even one phonon for about 97% of the time. They then fired a long train of pairs of laser light pulses (a few tens of millions) at the nanobeam, for which the intensity of the individual pulses was sufficiently low, and the intervals between the pulse pairs were sufficiently long,

to allow the nanobeam breathing mode to relax to its quantum ground state before the arrival of the next pulse pair.

The first pulse in a given pair was blue-detuned, enabling the pulse to 'write' a phonon into the nanobeam breathing mode, whereas the second pulse was red-detuned, potentially allowing it to 'read' a phonon out of the mode. The authors tuned the time interval between a given pulse pair, from a delay of one-tenth of a microsecond to a few microseconds, to determine how long the breathing mode could store a single phonon for. The scattered light produced from the nanobeam was split into two, and each half was directed to a different single-photon detector; each detector registered a voltage pulse only for a scattered photon that had written or read a phonon. Having two single-photon detectors enabled the detection of two scattered 'write' or 'read' photons resulting from the same light pulse. The probability of such detection events is extremely low because of the inherent weakness of the

photon force and the low photon-detection efficiencies, which is why so many pulse pairs were needed.

Riedinger and colleagues observed that the joint probability of detecting a scattered 'write' photon and a subsequent scattered 'read' photon significantly exceeds the joint probabilities of detecting two scattered 'write' photons or two scattered 'read' photons for the same pulse pair. This inequality, together with measurements of the latter two joint probabilities, provides strong evidence that the blue-detuned 'write' pulse puts the experimental system into a correlated quantum state, in which the number of photons in the cavity is always paired with the same number of phonons in the mechanical breathing mode. In particular, when the cavity is in the vacuum state (no scattered 'write' photons detected), then the mechanical mode must also be in the quantum ground state (no phonons). And for the rarer situation in which the cavity contains a single photon (one scattered 'write' photon detected), then the mechanical mode must also contain a single phonon. This photon–phonon pairing is the key result of the experiment.

The authors find that the correlated quantum states persisted for up to about 1 μ s, well short of the time taken for the vibrations of the breathing mode to dissipate (a few tens of microseconds). This might be because the nanobeam heats up during each 'write'-pulse stage. The storage lifetime of phonons might be lengthened by reducing the pulse intensities.

Having convincingly demonstrated a photon–phonon interface, a striking next step would be to generate an entangled quantum state that involves a single, breathing-mode phonon at the micrometre-wavelength scale on two silicon nanobeams separated from each other by up to 1 m or more. This could be achieved by bringing together 'write' photons scattered from both nanobeams and allowing them to interfere before being detected⁵. While the entangled state survives (possibly for up to a few microseconds), it would be meaningless to ascribe the phonon to one of the nanobeams and the vibrational ground state to the other. All that could be said is that the two nanobeams collectively possess the single phonon of vibrational energy — a bizarre state of affairs indeed. Such an entangled state has previously been demonstrated, but with much shorter-lived, higher-frequency (tens of terahertz) phonons⁶.

The ability to couple gigahertz and lower-frequency mechanical quantum-vibrational motion to other quantum systems (consisting of, for example, a few atoms, electrons or microwave photons) would allow nanomechanical resonators to serve as versatile interfaces that facilitate the transfer of quantum states between light and these other systems⁷. Together with the ability of light to transmit quantum states over large distances, this

would enable entanglement to be distributed between widely separated quantum systems — which would be useful for quantum information-processing applications. ■

Miles Blencowe is in the Department of Physics and Astronomy, 6127 Wilder Laboratory, Dartmouth College, Hanover, New Hampshire 03755, USA.
e-mail: miles.p.blencowe@dartmouth.edu

ANTHROPOLOGY

Hand of the gods in human civilization

Cross-cultural experiments find that belief in moralistic, knowledgeable and punishing gods promotes cooperation with strangers, supporting a role for religion in the expansion of human societies. SEE LETTER P.327

DOMINIC D. P. JOHNSON

In the modern world, we rely on governments, courts and the police to deter and punish those who would otherwise undermine social cooperation. But how did human societies achieve and sustain cooperation before these institutions existed? One possibility is religion: under the watchful gaze of supernatural agents, people modify their behaviour in an effort to avoid the wrath of the gods. In this issue, Purzycki *et al.*¹ (page 327)

1. Einstein, A., Podolsky, B. & Rosen, N. *Phys. Rev.* **47**, 777–780 (1935).
2. Riedinger, R. *et al.* *Nature* **530**, 313–316 (2016).
3. Aspelmeyer, M., Kippenberg, T. J. & Marquardt, F. *Rev. Mod. Phys.* **86**, 1391–1452 (2014).
4. Eichenfield, M., Camacho, R., Chan, J., Vahala, K. J. & Painter, O. *Nature* **459**, 550–555 (2009).
5. Duan, L.-M., Lukin, M. D., Cirac, J. I. & Zoller, P. *Nature* **414**, 413–418 (2001).
6. Lee, K. C. *et al.* *Science* **334**, 1253–1256 (2011).
7. Stannigel, K., Rabl, P., Sørensen, A. S., Zoller, P. & Lukin, M. D. *Phys. Rev. Lett.* **105**, 220501 (2010).

report a cross-cultural field-study finding that people are consistently more willing to give money to strangers of the same religion if the donor believes in a god that is moralizing (concerned about good and bad behaviour), knowledgeable (aware of one's thoughts and actions) and punishing (able to exact harm).

Pioneering anthropologists, such as Émile Durkheim and Bronisław Malinowski in the early twentieth century, have long argued that supernatural beliefs offer a powerful way to build materially cooperative societies. But in



Figure 1 | Weighing of the heart. This papyrus manuscript, a detail from the ancient Egyptian 'Book of the Dead' called *Papyrus of Ani*, depicts a scene in which the dead Ani's heart is weighed against a feather, representing Maat, goddess of truth and justice. At the top of the scene are the great Egyptian gods, ready to pronounce judgment on whether Ani should be granted entrance to the afterlife or banished to the underworld.



50 Years Ago

'The future of nuclear power' — The advanced type of gas-cooled reactor was expected in 1964 to cost less than a coal station of the same date, and within a year of the delivery of this Lecture the firm tender prices for the Dungeness B nuclear station showed that this was, in fact, the case, provided the station is built at the tender price. In the United States similar dramatic falls in costs have been experienced with their water moderated reactors; and Canada's heavy-water reactor is expected to have very low fuel costs, although it will have a high capital cost. These types of reactors, by the end of the century, would be using 100,000 tons of uranium per annum, on reasonable assumptions as to the rate of development of nuclear stations.

From *Nature* 19 February 1966

100 Years Ago

The memorandum regarding the neglect of science to which you refer in your leading article last week fails in my judgment by its moderation. The proposal that at least as many marks in the Civil Service examinations shall be allotted to science as to classics, may be a step in the right direction, but it is a halting one ... The revelations that have come to light in the course of this bloody war will, we hope, do at least this good, that the people may be induced to appreciate the necessity of basing education upon natural science instead of upon the classics. The appointment of a Minister of Science which is advocated in the memorandum would under existing conditions be of little use. Whatever qualifications he might be selected for, we may safely prophesy that entire ignorance of the subject he is to administer would be one.

From *Nature* 17 February 1916

the thriving new field of evolutionary religious studies, researchers are drawing on evolutionary theory to explore how religious beliefs can bring adaptive advantages — that is, contribute to an individual's survival or reproductive success. Although major debates remain², one theory that has gathered momentum is that a belief in supernatural punishment for violating social norms may be adaptive³ (Fig. 1).

How could this idea apply to cooperation? Deterring oneself from the pursuit of self-interest because of the risk of punishment from a watchful supernatural eye would seem to reduce an individual's evolutionary fitness, and should thus be eliminated by natural selection. However, even if such beliefs are false and costly, they may have generated net benefits: to individuals, by steering them away from selfish behaviour that risked retaliation in increasingly transparent and gossiping human societies; and/or to groups, by increasing the performance of the group as a whole in competition with other groups^{4,5}.

But what evidence do we have for such a theory? Empirical evidence that supernatural beliefs promote cooperation is mounting, but has tended to rely on qualitative, society-level or proxy measures of beliefs. Study participants have also typically been university students in developed nations, thus omitting the small-scale societies most relevant to the evolutionary problem at hand: how human groups achieved cooperation and made the transition from small to large societies in the first place. Perhaps the most important lacuna is that previous studies have not rigorously addressed whether the beliefs of the recipients of cooperative acts changes people's generosity towards them.

Purzycki and colleagues' study addresses many of these issues by using controlled experimental games among participants from eight small-scale societies around the world and tying the results to explicit measures of individuals' beliefs. Participants played a simple but clever game (designed to subtly reveal preferences), in which they allocated coins between a distant co-religionist (people who were members of the same religion, but who lived geographically far away) and either themselves or a local co-religionist. The researchers found that the more subjects rated their god as moralistic, knowledgeable and punishing, the more money they gave to distant strangers adhering to the same religion. Notably, belief in rewards from the god could not account for the results — supernatural punishment seemed responsible.

Because the study is correlational, one worry is that some unexamined variable could account for the results — perhaps certain people are disposed to both kindness to strangers and belief in punitive gods, for example. However, Purzycki *et al.* show that allocations increased for moralistic gods that were punishing and knowledgeable, but not

for more locally relevant supernatural agents that were also punishing and knowledgeable. Hence, general conceptions of supernatural agents cannot alone explain the results. Rather, it is moralistic, 'big' gods that seem to stimulate generosity towards distant co-religionists⁶.

The authors did not conduct experiments to assess allocations to oneself versus a local co-religionist, nor experiments involving non-religious recipients, so we don't know whether local supernatural agents might promote cooperation between individuals within the local community, as other work has found⁷, or whether any kind of god promotes cooperation with strangers of another, or no, religion. Purzycki *et al.* focused on cooperation with co-religionists beyond the local community, and thus the expansion of human society from small to large groups. But future studies of the role of local gods are needed to improve our understanding of the evolutionary origins of religion (before there were big groups or big gods), and of whether and how religion brings adaptive advantages to individuals⁸.

It is worth emphasizing that the subjects in this experiment were not cooperative with random strangers, only with strangers that shared the same god. We therefore still face the challenge of understanding the promotion of cooperation and trust among members of different religions. Purzycki and colleagues' finding that sharing the same god is key to cooperation suggests that this may be an even harder nut to crack. In fact, one of the most compelling explanations for why individuals may help the group at their own expense is that it aids survival in an environment of inter-group competition. Whenever the threat of exploitation or warfare is present, the best protection is larger and more-cohesive societies, which are better able to deter or defeat rivals. Religion's positive role in reducing self-interest and promoting cooperation may therefore reflect the costs of competition as much as the benefits of generosity⁹.

Religion is arguably the most powerful mechanism that societies have found to bind people together in common purpose. From ancient civilizations, to the spread of Christianity, to today's Islamist terrorist groups, religion has motivated not only the subordination of self-interest for the wider group, but even martyrdom in the name of a god. We are still grappling to understand, from a scientific perspective, why and under what circumstances humans sacrifice their own welfare for the benefit of distant others¹⁰. But there is little doubt about the power of religion to promote allegiance to one's god and group. Purzycki and colleagues' study offers the most explicit evidence yet that belief in supernatural punishment has been instrumental in boosting cooperation in human societies. A large part of the success of human civilizations may have lain in the hands of the gods, whether or not they are real. ■

Dominic D. P. Johnson is in the Department of Politics and International Relations, University of Oxford, Oxford OX1 3UQ, UK. e-mail: dominic.johnson@politics.oxford.ac.uk

1. Purzycki, B. G. et al. *Nature* **530**, 327–330 (2016).
2. Sosis, R. J. *Cogn. Culture* **9**, 315–332 (2009).
3. Johnson, D. *God is Watching You: How the Fear of God Makes Us Human* (Oxford Univ. Press, 2016).

4. Schloss, J. P. & Murray, M. J. *Religion Brain Behav.* **1**, 46–99 (2011).
5. Wilson, D. S. *Darwin's Cathedral: Evolution, Religion, and the Nature of Society* (Univ. Chicago Press, 2002).
6. Norenzayan, A. *Big Gods: How Religion Transformed Cooperation and Conflict* (Princeton Univ. Press, 2013).
7. Hartberg, Y., Cox, M. & Villamayor-Tomas, S. *Religion Brain Behav.* <http://dx.doi.org/10.1080/2153599X.2014.959547> (2014).
8. Watts, J. et al. *Proc. R. Soc. B* **282**, 20142556 (2015).
9. Turchin, P. *Ultrasociety: How 10,000 Years of War Made Humans the Greatest Cooperators on Earth* (Beresta, 2015).
10. Whitehouse, H. & Lanman, J. A. *Curr. Anthropol.* **55**, 674–695 (2014).

This article was published online on 10 February 2016.

CLIMATE SCIENCE

Ice streams waned as ice sheets shrank

It emerges that ice discharge from a major ice sheet did not increase rapidly at the end of the most recent ice age. The finding points to steady, not catastrophic, ice-sheet loss and sea-level rise on millennial timescales. **SEE LETTER P.322**

JASON P. BRINER

Ice sheets and the global sea are locked in a tug of war for water in which climate change dictates which side gains or loses ground. Mindful that tug-of-war contests often end with the catastrophic collapse of one side, climate scientists are deeply concerned about the manner in which ice sheets are currently declining¹. This concern is largely focused on mechanisms that amplify ice loss, such as the acceleration of ice streams — massive rivers of ice that drain a disproportionately large amount of ice from ice sheets² (Fig. 1). On page 322 of this issue, Stokes *et al.*³ present the first complete reconstruction of ice-stream activity throughout the disintegration of an ice sheet.

A fleet of geoscientists scattered across the poles, along with satellites in space, watches every move that ice streams make in Greenland and the Antarctic⁴. This work matters because ice streams are the valves that control the enormous volumes of ice poised to spill into the oceans and cause a rapid jump in sea level. But the duration of these observations (tens of years) has been short compared with the time that it takes for substantial ice-sheet changes to occur (hundreds to thousands of years).

One way to gauge ice-stream activity on longer timescales is to use the geological record — in sediments and landforms — of past ice streams that left their imprint on formerly glaciated landscapes. Over the past two decades, Stokes and his co-workers have greatly advanced the methods used to assemble scraps of evidence left behind by extinct ice streams and to generate histories of ice-stream activity⁵. The present study builds on this foundation and on the established chronology of ice-sheet positions through time⁶.

Stokes *et al.* evaluate the importance of ice streams at the end of the most recent ice age, when a tug of war also played out between ice sheets and the ocean. Their evidence shows that ice streams turned on and off, and shifted

from place to place, during the disappearance of the Laurentide Ice Sheet — the Antarctic-sized ice sheet that occupied Canada and the northern United States at that time. Perhaps most notably, Stokes and colleagues find that ice-stream activity decreased as the planet warmed: the number of ice streams fell, the amount of ice expunged by them decreased and ice streams occupied a progressively smaller percentage of the ice-sheet edge.

The authors' findings represent a leap forward in our view of ice-stream activity on timescales longer than a few decades. Until now, we were in the dark about how ice streams respond to ice-sheet decay. Would excessive ice streaming lower the elevation of ice sheets, thus robbing ice-accumulation centres of their elevated positions (which are good for gathering snow that compresses to form ice), and



Figure 1 | An Antarctic ice stream. Fast-moving rivers of ice, such as the Byrd Glacier (pictured) are known as ice streams, and drain a disproportionately large amount of ice from ice sheets. Stokes *et al.*³ report a complete reconstruction of ice-stream activity throughout the disintegration of the ancient Laurentide Ice Sheet.

triggering catastrophic ice-sheet collapse? Does the additional meltwater that forms during periods of warming lubricate ice streams, causing them to discharge ice faster for prolonged periods? Stokes and co-workers' results suggest not.

However, the relevance of these findings to future ice-sheet behaviour is not totally clear, because the Laurentide Ice Sheet is not an exact analogue of today's ice sheets. For example, much of the Laurentide (including the ice streams embedded in it) terminated on land, whereas ice streams within the Greenland and Antarctic ice sheets terminate in the sea. Furthermore, present-day ice streams are largely 'fixed' in space by the mountains through which they pass⁷, and therefore could flow for thousands of years. This differs from the many Laurentide ice streams that were not confined by the underlying landscape, and thus were typically more ephemeral.

Predicting the pace of future ice loss and sea-level rise is an enormous challenge. With their improved sophistication, numerical ice-sheet models have led to great strides in our understanding of how quickly ice sheets may vanish in the future, particularly marine-based ice sheets such as the West Antarctic Ice Sheet^{8,9}. Stokes and co-workers' strategy of extracting data from relict ice-age landscapes provides a new viewpoint. As the Greenland Ice Sheet, and particularly the East Antarctic Ice Sheet, eventually retreat from the ocean, the Laurentide Ice Sheet becomes more closely analogous to them. Continued effort is needed to discover other secrets hidden in the palaeo-record of ice-sheet response to past climate change.

The current findings may not provide guidance about ice-stream changes during this century, but they will help us to predict the pace of long-term ice-sheet disappearance. And although ice sheets will inevitably lose the current tug of war, it is reassuring to know that the competition might not end catastrophically. ■

Jason P. Briner is in the Department of Geological Sciences, State University of New York at Buffalo, New York 14260, USA. e-mail: jbriner@buffalo.edu

- Church, J. A. et al. in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. et al.) 1137–1216 (Cambridge Univ. Press, 2013).
- Bennett, M. R. *Earth-Sci. Rev.* **61**, 309–339 (2003).
- Stokes, C. R., Margold, M., Clark, C. D. & Tarasov, L. *Nature* **530**, 322–326 (2016).
- Joughin, I., Smith, B. E., Shean, D. E. & Floricioiu, D. *Cryosphere* **8**, 209–214 (2014).
- Stokes, C. R. & Clark, C. D. *Ann. Glaciol.* **28**, 67–74 (1999).
- Dyke, A. S., Moore, A. & Robertson, L. *Geol. Surv. Can. Open File Rep.* 1574 (2003).
- Truffer, M. & Echelmeyer, K. A. *Ann. Glaciol.* **36**, 66–72 (2003).
- Golledge, N. R. et al. *Nature* **526**, 421–425 (2015).
- Pollard, D., DeConto, R. M. & Alley, R. B. *Earth Planet. Sci. Lett.* **412**, 112–121 (2015).

CELL BIOLOGY

Form follows function for mitochondria

The fission of organelles called mitochondria has now been linked to the stress-sensor protein AMPK. When activated by stress, this protein phosphorylates the mitochondrial receptor protein MFF, which recruits the fission machinery.

CHUNXIN WANG & RICHARD YOULE

The architect Louis Sullivan wrote that, for all things, "form ever follows function". In keeping with this rule, energy-producing organelles called mitochondria exist in different forms in different cell types^{1,2} and under different conditions³. Long, contiguous assemblies of mitochondria promote energy production, whereas stress causes their fragmentation into small, round, disconnected units. Mitochondria switch between these two forms through rounds of fission and fusion⁴. Writing in *Science*, Toyama *et al.*⁵ identify a signalling pathway that triggers the fission of long mitochondrial assemblies in response to stress.

The energy status of a cell can be measured by the ratio of AMP molecules (an end product of chemical-energy expenditure) to energy-carrying ATP molecules in the cytoplasm. Mitochondria produce ATP in response to energy depletion. One sensor of energy depletion is an evolutionarily conserved protein known as AMP-activated protein kinase (AMPK). This protein is activated by binding AMP or by stresses that deplete ATP and increase the AMP:ATP ratio, such as glucose deprivation, inadequate blood supply and lack of oxygen⁶.

Activation of AMPK enhances the production of mitochondria and improves endurance — indeed, the use of drugs that activate AMPK, such as AICAR, is banned from competitive sports by the World Anti-Doping Agency. AMPK phosphorylates several substrates⁶, thereby regulating metabolism, mitochondrial proliferation and an intracellular degradation process called autophagy. However, previous studies have not demonstrated whether AMPK directly regulates mitochondrial form.

Researchers have known for years⁷ that mitochondria undergo fission and become fragmented in response to poisons such as cyanide, which damage the organelles by inhibiting the electron transport chain (ETC) that drives ATP synthesis. But how such mitochondrial dysfunction is sensed, and how it triggers a rapid fission response, has remained unclear.

Toyama *et al.* found that AMPK is rapidly activated in response to the ETC-inhibiting drugs rotenone and antimycin A, presumably

as a consequence of an increased AMP:ATP ratio. Mitochondria became fragmented in drug-treated cells, but this response was prevented by deletion of the gene that encodes AMPK. By contrast, pharmacological activation of AMPK with AICAR was sufficient

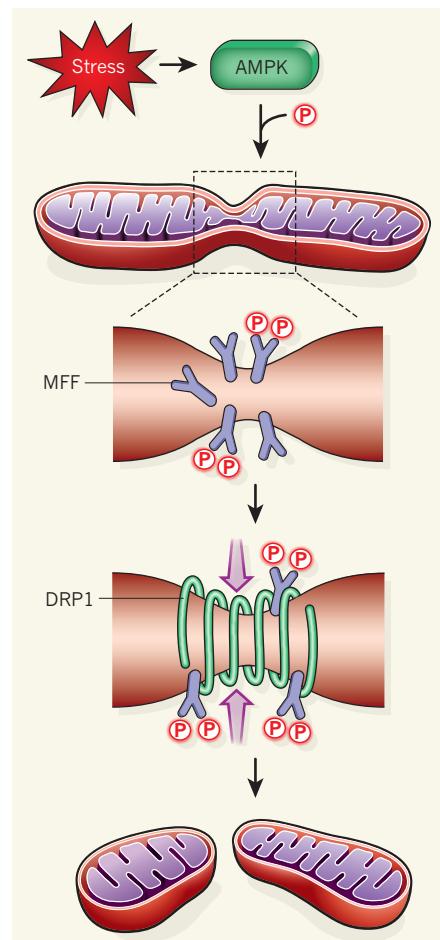


Figure 1 | The stress of separation. Mitochondria, which are the cell's energy centres, sometimes exist in fused assemblies that can be fragmented in response to stress. The receptor protein MFF is located on the outer membrane of such assemblies, at putative separation sites. Toyama *et al.*⁵ report that, when activated in response to stress, the enzyme AMP-activated protein kinase (AMPK) phosphorylates (P) MFF, which then recruits the protein Dynamin-related protein 1 (DRP1) to the membrane. DRP1 forms constricting spiral complexes around the mitochondria, mediating fission.

to cause fragmentation, demonstrating a surprising link between activation of this sensor protein and mitochondrial fission.

How does AMPK relay signals to the mitochondrial fission machinery? The core mitochondrial fission factor is the enzyme Dynamin-related protein 1 (DRP1), which wraps around constriction sites at the mitochondrial membrane, forming spiral complexes that mediate scission. DRP1 is primarily located in the cytoplasm^{8,9}, but is recruited to prospective mitochondrial scission sites by receptor proteins such as MFF (ref. 10). Toyama and colleagues showed that AMPK directly phosphorylates MFF at two amino-acid residues, serine-155 (S155) and S172, consistent with results from a previous study¹¹. This modification increased the recruitment of DRP1 to mitochondria (Fig. 1). Furthermore, mutations that blocked MFF phosphorylation prevented poison-mediated mitochondrial fragmentation and DRP1 recruitment.

Thus, the authors have identified a mechanism that regulates mitochondrial form, and probably function. They propose that the increased fission induced by AMPK facilitates selective elimination of the damaged mitochondria through autophagy, which may act as a quality-control response to the poisons. In support of this suggestion, mitochondrial fission has previously been linked to engulfment and elimination of damaged mitochondria by autophagosome structures, which mediate autophagy¹². Therefore, the authors' proposal is a plausible explanation for their observations, and deserves further study.

Although glucose deprivation or starvation can also increase the AMP:ATP ratio and activate AMPK, starvation does not lead to fission, but actually inhibits the process, causing mitochondria to elongate^{13,14}. How different stimuli that activate AMPK can lead to distinct mitochondrial responses will be an interesting avenue for exploration. Perhaps, for instance, differential modulation of DRP1 phosphorylation controls these varied responses. There is evidence^{13,14} that starvation of cells causes DRP1 phosphorylation at S637. It has also been reported¹⁵ that AMPK activation indirectly phosphorylates DRP1 at S637, and this phosphorylation has been linked to the inhibition of DRP1 and a decrease in mitochondrial fission¹⁶.

Although it is not clear how the activation of AMPK by different energetic stressors yields opposing effects on mitochondrial morphology, it seems apparent that damage is distinguished from energetic demand, leading to fission and quality control, rather than fusion and energy production. The pathway identified by Toyama *et al.* highlights the importance of fine-tuning mitochondrial-fission responses to different stimuli. Because form follows function, the authors' study implies that AMPK works in the context of other cellular signals to promote a variety of mitochondrial functions. ■

Chunxin Wang and Richard Youle are in the Biochemistry Section, Surgical Neurology Branch, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland 20892, USA. e-mails: wangchu@ninds.nih.gov; youler@ninds.nih.gov

- Glancy, B. *et al.* *Nature* **523**, 617–620 (2015).
- Das, S. *et al.* *Pflügers Arch.* **464**, 101–109 (2012).
- Rossignol, R. *et al.* *Cancer Res.* **64**, 985–993 (2004).
- Chan, D. C. *Annu. Rev. Genet.* **46**, 265–287 (2012).
- Toyama, E. Q. *et al.* *Science* **351**, 275–281 (2016).
- Hardie, D. G. *Genes Dev.* **25**, 1895–1908 (2011).
- Youle, R. J. & van der Bliek, A. M. *Science* **337**, 1062–1065 (2012).
- Otsuga, D. *et al.* *J. Cell Biol.* **143**, 333–349 (1998).
- Smirnova, E., Griparic, L., Shurland, D.-L. & van der Bliek, A. M. *Mol. Biol. Cell* **12**, 2245–2256 (2001).
- Otera, H. *et al.* *J. Cell Biol.* **191**, 1141–1158 (2010).
- Ducommun, S. *et al.* *Cell Signal.* **27**, 978–988 (2015).
- Twig, G. *et al.* *EMBO J.* **27**, 433–446 (2008).
- Rambold, A. S., Kostelecky, B., Elia, N. & Lippincott-Schwartz, J. *Proc. Natl. Acad. Sci. USA* **108**, 10190–10195 (2011).
- Gomes, L. C., Benedetto, G. D. & Scorrano, L. *Nature Cell Biol.* **13**, 589–598 (2011).
- Wikstrom, J. D. *et al.* *Mol. Endocrinol.* **27**, 1706–1723 (2013).
- Chang, C.-R. & Blackstone, C. J. *Biol. Chem.* **282**, 21583–21587 (2007).

DEVELOPMENTAL BIOLOGY

Females have a lot of guts

The discovery of sex-biased proliferation in the intestinal stem cells of fruit-fly midguts reveals that the organ's size is determined by a previously undefined, sex-specific molecular pathway. SEE LETTER P.344

JUSTIN FEAR & BRIAN OLIVER

Within-species genetic differences are central to many biological phenomena, from evolution to disease susceptibility. One of the most under-studied intraspecies differences is sex. We tend to think of sex in terms of characteristics that are related to reproduction, but differences between sexes extend to many parts of the body. For instance, the midgut (an absorptive organ similar to the small intestine) of the fruit fly *Drosophila melanogaster* is longer in females than in males — especially after mating, when females produce many eggs that are replete with proteins and lipids¹. On page 344 of this issue, Hudry *et al.*² report that a previously unidentified branch of the sex-determination pathway underlies this dynamic difference in organ size, by controlling the proliferation of stem cells in the midgut.

The classic view of sex determination in fruit flies involves a regulatory cascade in which genes are spliced into different forms in a sex-specific manner³. These early steps in fly sex determination differ from those in the mammalian set-up, but the ultimate outcomes are similar. In flies with two X chromosomes, the protein Sex lethal (Sxl) splices an RNA called *transformer* (*tra*) into a protein-coding isoform. The *TRA* protein, acting with its cofactor *TRA2*, binds to both the RNA produced from the *doublesex* (*dsx*) gene, splicing it into a female-specific isoform, and to the RNA of *fruitless* (*fru*), blocking production of a male-specific isoform. In XY flies, the gene *Sxl* is not expressed. As such, *dsx* and *fru* are spliced into

male-specific isoforms by default. Through another pathway, *Sxl* prevents dosage compensation in females (in males, transcription of genes on the X chromosome is upregulated, to compensate for the fact that females have two copies of X).

Hudry *et al.* investigate sex-specific differences in gene expression in the fruit-fly midgut. They report that genes involved in cell division are preferentially expressed in females. Furthermore, they find that *Sxl* acts to enhance the proliferative capacity of intestinal stem cells in the female midgut. These differences from the male midgut help to explain the larger, more-plastic guts found in females.

The authors go to great lengths to demonstrate that these sex-specific differences are not regulated by the classic splicing pathway. They rule out a role for *dsx* and *fru* in female gut growth and plasticity and provide evidence that *TRA2* might also be dispensable. However, there is a caveat to this suggestion — although mutation of *tra2* had no effect on proliferation, there was still *TRA2* activity in these flies. By contrast, *tra* expression is required for proliferation to be enhanced in females. Thus, sex differences in the fruit-fly midgut are regulated by a previously unidentified branch of the sex-determination pathway, one that is downstream of *Sxl* and *TRA*.

Plasticity in the female gut is blocked by downregulating *Sxl*, but Hudry and colleagues show that plasticity can be rescued by the expression of *TRA*. These data make a convincing case that sex differences in the gut are mediated by *tra* rather than by dosage compensation. Moreover, misexpression of *tra* in

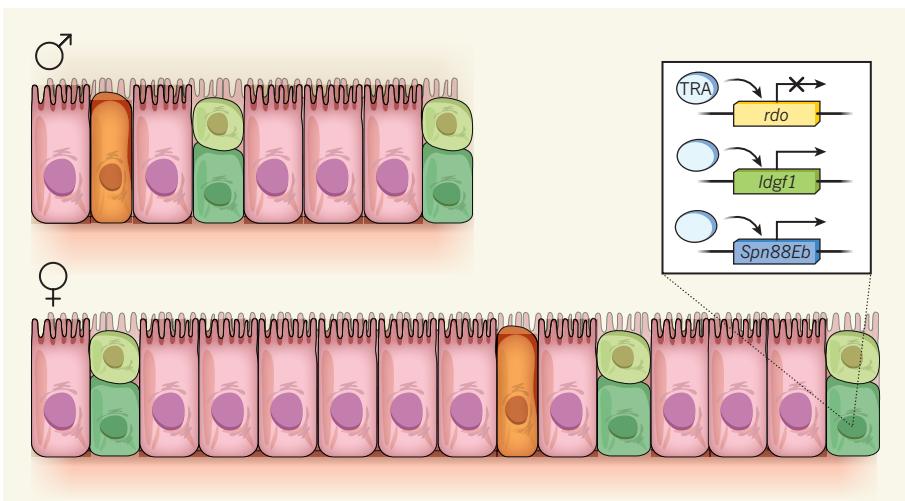


Figure 1 | Sex and the gut. The midgut of female fruit flies (*Drosophila melanogaster*) is longer than that of males. It contains intestinal stem cells (dark green), which give rise to intestinal progenitors (light green), which then differentiate into the mature cell types of the intestine (pink and orange). Hudry *et al.*² report that proliferation of intestinal stem cells is enhanced in female midguts compared with that in males. They find that the protein TRA, which is produced only in females, promotes this sex-specific proliferation by modulating expression of the genes *reduced ocelli* (*rdo*), *Imaginal disc growth factor 1* (*Idgf1*), *Serpin 88Eb* (*Spn88Eb*) and perhaps others.

intestinal progenitors enhances proliferation in the male midgut. Thus, unlike differentiated sex-specific organs such as the male accessory gland (the equivalent of the human prostate), the sex effects on intestinal stem cells are fully reversible.

Next, the authors look for direct and indirect targets of TRA in intestinal stem cells by expression profiling. They find 72 genes for which RNA splicing or steady-state expression levels are modulated by TRA, three of which encode proteins that regulate proliferation of the cells (Fig. 1). TRA is thought of as an RNA-binding protein and splicing factor⁴, so direct targets would be expected to be regulated post-transcriptionally. However, in these three targets, TRA modulates transcript abundance. It is possible that these genes are indirect targets of TRA and are regulated by an unidentified downstream transcription factor. Alternatively, TRA might have a direct effect on transcript stability, or even on the regulation of transcription.

Cell-autonomous events (those that occur on a cell-by-cell basis, rather than affecting neighbouring cells) have received much attention in sex-determination research in *D. melanogaster*, whereas vertebrate work has focused mainly on the influence of hormones such as testosterone. However, coordinating sex-biased expression between cells and organs in fruit flies clearly requires broader control. Indeed, although *tra* is required only in the intestinal stem cells in which it is expressed, two of its three proliferation-regulating targets — *Imaginal disc growth factor 1* (*Idgf1*) and *Serpin 88Eb* (*Spn88Eb*) — encode secreted proteins. *Idgf* proteins are known to increase proliferation⁵ and might bind to receptor

proteins on the intestinal stem cells from which they are secreted to promote proliferation directly. Serpin proteins can also bind to a variety of protein types, including hormones, which they may escort around the body⁶.

In addition to having longer guts, female fruit flies are larger than males. Recent work⁷ shows that fat-cell expression of *tra*, but not of *dsx* or *fru*, contributes autonomously to cell size and systemically to female body size, by promoting insulin secretion from the brain. *Idgf1* and *Spn88Eb* are highly expressed in fat cells⁸ — could they be TRA targets there, too? Of note, the protein Serpinb1 mediates plasticity in insulin-producing cells in mice⁹, suggesting that this pathway might be evolutionarily conserved. There is still much to learn, but the sex-biased sizing of organisms and organs seems likely to be under the control of an entirely

unknown branch of the sex-determination cascade, with both cell-autonomous and non-cell-autonomous functions.

In a final series of experiments, Hudry and colleagues demonstrate that sex-specific organ plasticity alters susceptibility to disease, because female flies are more sensitive than males to tumours that are induced through genetic changes. Little is known about sex-biased disease susceptibility in *D. melanogaster*, although sex-specific interactions are implicated in a host of disease-related traits¹⁰, suggesting that sex is an important consideration for understanding disease in flies. Sex bias is seen in many human diseases¹¹, including several digestive-tract cancers. Furthermore, although the authors show substantial sex-biased gene expression in the fruit-fly gut, many of the studies on digestive-tract tumour development in *D. melanogaster* looked only at females. Thus, the current study highlights how influences on disease can be missed by ignoring sex chromosomes and sexual identity, even in non-mammalian model organisms. ■

Justin Fear and Brian Oliver are in the Laboratory of Cellular and Developmental Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA.

e-mail: briano@helix.nih.gov

1. Reiff, T. *et al.* *eLife* **4**, e06930 (2015).
2. Hudry, B., Khadate, S. & Miguel-Aliaga, I. *Nature* **530**, 344–348 (2016).
3. Clough, E. & Oliver, B. *Brief. Funct. Genomics* **11**, 387–394 (2012).
4. Lynch, K. W. & Maniatis, T. *Genes Dev.* **10**, 2089–2101 (1996).
5. Bryant, P. J. *Novartis Found. Symp.* **237**, 182–202 (2001).
6. Silverman, G. A. *et al.* *J. Biol. Chem.* **285**, 24299–24305 (2010).
7. Rideout, E. J., Narsaiya, M. S. & Grewal, S. S. *PLoS Genet.* **11**, e1005683 (2015).
8. Brown, J. B. *et al.* *Nature* **512**, 393–399 (2014).
9. El Ouamari, A. *et al.* *Cell Metab.* **23**, 194–205 (2016).
10. Mackay, T. F. C. *Phil. Trans. R. Soc. Lond.* **365**, 1229–1239 (2010).
11. Arnold, A. P. *Biol. Sex Differ.* **1**, 1 (2010).

GENOMICS

From sea to sea

The genome sequence of the marine flowering plant eelgrass (*Zostera marina*) sheds light on how marine algae evolved into land plants before moving back to the sea. **SEE LETTER P.331**

SUSAN L. WILLIAMS

Eelgrass (*Zostera marina*) is a member of a family of highly specialized, sexually reproducing, marine flowering plants known collectively as seagrasses. It is an unlikely model for plant evolution, but is

a useful one because it has undergone major habitat shifts: it evolved from marine algae into a terrestrial flowering plant, then moved back to the sea again. In this issue, Olsen *et al.*¹ (page 331) describe the complete genome sequence of eelgrass. Their study marks the culmination of 8 years' work by 35 scientists

from around the world, and should help plant biologists not only to dissect how eelgrass evolved, but also to gain a better understanding of flowering-plant evolution in general.

Seagrasses have been largely ignored by evolutionary biologists, perhaps because they are so often out of sight. But eelgrass can form vast green coastal-sea meadows that support many species (Fig. 1), including ecologically and economically valuable organisms such as halibut, clams and endangered sea otters. As such, it is a good model system in which to analyse how genetics plays into ecosystem functioning — an understanding of which is the linchpin for human efforts to conserve biological diversity in the face of a rapidly changing Earth. Moreover, eelgrass is as productive in terms of biomass as are maize (corn) and sugar cane, and its root mat stabilizes sediments and thus shorelines. Indigenous peoples from the US Pacific Northwest to Mexico have built cultures around eelgrass, and Europeans have used it to stuff furniture and to insulate homes, and have even grazed cattle on intertidal meadows².

Eelgrass is unusual from an evolutionary standpoint, too. The plant completed a remarkable feat when it readapted from a freshwater to a marine lifestyle and became able to compete with seaweeds. Its evolutionary path began with marine green algae, which evolved to cope with terrestrial habitats and to produce flowers and seeds. Then, angiosperms (the collective name for flowering plants) entered fresh water — an evolutionary step that seems to have been made many times. From the freshwater angiosperm lineage, eelgrass, along with a few other seagrass species, diverged to return to the sea. But it seems that this step occurred just three times³, indicating the extreme nature of this habitat shift.

Despite the plant's evolutionary and ecological importance, genetic analysis of eelgrass did not begin until the 1990s (ref. 4), more recently than for other angiosperms such as wheat, peas and land-dwelling grasses. The early studies were fraught with setbacks, because it proved difficult to purify the DNA and protein variants found in different plants of the species. Furthermore, scientists have never successfully artificially selected or genetically engineered eelgrass. Thus, the classic approach of crossing plants cannot be used to delve into seagrass genetics. Given this problematic history, Olsen and colleagues' sequence represents a major advance.

The complete genome sequence reveals that, in moving from calm lakes and ponds to the rough, salty ocean, eelgrass lost several key gene groups. It lost all of the genes involved



Figure 1 | Eelgrass ecosystems. The eelgrass *Zostera marina* is home to an abundance of wildlife, including sea anemones.

in stomata (pores on plant leaves that regulate gas exchange and minimize water loss). These pores are not essential in seagrasses, because the plants are not prone to moisture loss, and they absorb dissolved gases directly through outer cell layers. The organism also lost genes that confer protection from ultraviolet light, as well as those involved in sensing far-red light — these wavelengths do not penetrate very far in coastal waters.

During the move to the sea, eelgrass regained genes encoding cell-wall compounds that were lost when marine algae transitioned to land. These genes have crucial roles in allowing osmotic adjustment to salt, and in promoting nutrient uptake and gas exchange in saltwater environments. Other evolutionary innovations include changes that enable pollen to stick to stigmas (the tips of female flower parts) in salt water, an expanded capacity to capture light and to photosynthesize in dim coastal seas, and a loss of genes encoding the proteins that synthesize and sense volatile terpenes, compounds that are common in aromatic herbs but that would not be effective in an aqueous environment in their putative role of deterring predators in the ocean.

The eelgrass genome is valuable for several reasons. For evolutionary biologists, it represents a missing piece in the puzzle of angiosperm evolution. It also provides a wealth of information that will improve our

understanding of diverse biochemical pathways. For example, identification of the DNA sequences of genes that confer tolerance to salt water means that the plant could provide a model system in which to study how agricultural crops might adjust to increasingly saline soils.

Eelgrass is remarkably adaptable, growing under ice in the Arctic Ocean and surviving scorching heat in the Mexican state of Baja California. It also has the largest distribution of any plant in the temperate Northern Hemisphere. For marine ecologists, the genome is a powerful tool for uncovering the adaptations that allow the plant to thrive in a wide range of environmental conditions. This ability to adapt might be the key to surviving environmental changes such as ocean acidification, warming and freshening that are occurring under global climate change. It is already known that eelgrass populations that are more genetically diverse survive disturbances better, can be restored more rapidly, produce more biomass and support more animals than their less-diverse counterparts⁵. Now, the genome will help researchers to delve into exactly which genetic elements facilitate such high biomass production and resilience.

Olsen and colleagues' genome-sequencing feat may have come just in time. Seagrass ecosystems are being lost rapidly, with seagrass fields disappearing at a global rate of about one American-football field every 30 minutes; some species and their associated animals are even threatened by extinction⁶. The disappearances are attributable to human disturbances such as building marinas, over-fertilizing coastal waters and aquaculture⁷. Seagrass restoration and conservation efforts are under way across the globe, and an understanding of the genes that adapt these fascinating species to marine life can only help these efforts. ■

Susan L. Williams is at the Bodega Marine Laboratory and Department of Evolution and Ecology, University of California, Davis, Bodega Bay, California 94923, USA.
e-mail: slwilliams@ucdavis.edu

1. Olsen, J. L. *et al.* *Nature* **530**, 331–335 (2016).
2. Hemminga, M. A. & Duarte, C. M. *Seagrass Ecology* (Cambridge Univ. Press, 2000).
3. Les, D. H., Cleland, M. A. & Waycott, M. *Syst. Bot.* **22**, 443–463 (1997).
4. Ruckelshaus, M. H. *Mar. Biol.* **123**, 583–593 (1995).
5. Duffy, J. E. *et al.* *Ecol. Lett.* **18**, 696–705 (2015).
6. Hughes, A. R., Williams, S. L., Duarte, C. M., Heck, K. L. Jr & Waycott, M. *Front. Ecol. Environ.* **7**, 242–246 (2009).
7. Waycott, M. *et al.* *Proc. Natl. Acad. Sci. USA* **106**, 12377–12381 (2009).

This article was published online on 27 January 2016.

The peptidergic control circuit for sighing

Peng Li^{1*}, Wiktor A. Janczewski^{2*}, Kevin Yackle^{1*}, Kaiwen Kam^{2†}, Silvia Pagliardini^{2†}, Mark A. Krasnow^{1§} & Jack L. Feldman^{2§}

Sighs are long, deep breaths expressing sadness, relief or exhaustion. Sighs also occur spontaneously every few minutes to reinflate alveoli, and sighing increases under hypoxia, stress, and certain psychiatric conditions. Here we use molecular, genetic, and pharmacologic approaches to identify a peptidergic sigh control circuit in murine brain. Small neural subpopulations in a key breathing control centre, the retrotrapezoid nucleus/parafacial respiratory group (RTN/pFRG), express bombesin-like neuropeptide genes neuromedin B (*Nmb*) or gastrin-releasing peptide (*Grp*). These project to the preBötzi Complex (preBötC), the respiratory rhythm generator, which expresses NMB and GRP receptors in overlapping subsets of ~200 neurons. Introducing either neuropeptide into preBötC or onto preBötC slices, induced sighing or *in vitro* sigh activity, whereas elimination or inhibition of either receptor reduced basal sighing, and inhibition of both abolished it. Ablating receptor-expressing neurons eliminated basal and hypoxia-induced sighing, but left breathing otherwise intact initially. We propose that these overlapping peptidergic pathways comprise the core of a sigh control circuit that integrates physiological and perhaps emotional input to transform normal breaths into sighs.

A sigh is a long, deep breath often associated with sadness, yearning, exhaustion or relief. Sighs also occur spontaneously, from several per hour in humans to dozens per hour in rodents^{1,2}. Their recurrence during normal breathing enhances gas exchange and may preserve lung integrity by reinflating collapsed alveoli^{2–4}. Sighing increases in response to emotional and physiological stresses, including hypoxia and hypercapnia, and in anxiety disorders and other psychiatric conditions where it can become debilitating^{5,6}.

The core of the breathing rhythm generator is the preBötC, a cluster of several thousand neurons in ventrolateral medulla. preBötC is required for inspiration and generates respiratory rhythms in explanted brain slices^{7–10}. Each rhythmic burst activates premotoneurons and motoneurons that contract the diaphragm and other inspiratory muscles, generating a normal ('eupneic') breath¹⁰. Occasionally, a second preBötC burst immediately follows the first, and this 'double burst' leads to the augmented inspiration of a sigh^{11–13}, typically about twice the volume of a normal breath¹⁴. Thus, the command for normal breaths and sighs both appear to emanate from preBötC.

A variety of neuromodulators and neuropeptides^{15,16}, including frog bombesin¹⁵, can influence sighing in rodents. However, the endogenous sigh control pathways have not been identified. We tested the effect of injecting bombesin into preBötC¹⁷, and we screened for genes selectively expressed in breathing control centres (K.Y. and M.A.K., in preparation). These parallel approaches led to identification of two bombesin-like neuropeptide pathways connecting the RTN/pFRG, another medullary breathing control centre^{18,19}, to preBötC. We provide genetic, pharmacologic and neural ablation evidence that these pathways are critical endogenous regulators of sighing, and define the core of a dedicated sigh control circuit.

Neuromedin B links two breathing control centres

To identify breathing control genes, we screened >19,000 gene expression patterns in embryonic day 14.5 mouse hindbrain²⁰ (K.Y. and M.A.K., in preparation). The most specific pattern was *Nmb*, one of two genes encoding bombesin-like neuropeptides in mammals. *Nmb*

is expressed in the medulla surrounding the lateral half of the facial nucleus, in or near RTN/pFRG in mouse (Fig. 1a, b and Extended Data Fig. 1a) and rat (Extended Data Fig. 1b). *Nmb* mRNA was also detected in the olfactory bulb and hippocampus (Extended Data Fig. 1c, d).

Nmb expression was further characterized using an *Nmb*-GFP BAC transgene with the *Nmb* promotor driving GFP expression, which reproduced the endogenous *Nmb* pattern (Fig. 1b, c). *Nmb*-GFP expressed in 206 ± 21 (mean \pm standard deviation (s.d.), $n = 4$) RTN/pFRG neurons per side, most of which (92%, $n = 53$ cells scored) co-expressed *Nmb* mRNA (Extended Data Fig. 1e–h). In CLARITY-processed²¹ brainstems, GFP-labelled cells surrounded the lateral half of the facial nucleus, with the highest density ventral and dorsal (Fig. 1d, e and Supplementary Video 1). This ventral parafacial region is the RTN, an important sensory integration centre for breathing^{18,22,23}. Nearly all *Nmb*-GFP-positive cells (96%; $n = 202$ cells from 2 animals) co-expressed canonical RTN marker PHOX2B²⁴ (Fig. 1f), comprising one-quarter of the ~800 PHOX2B-positive RTN neurons²⁵.

Nmb-expressing neurons projected to preBötC (Fig. 1g, j). Punctate NMB staining was detected along the projections (Extended Data Fig. 2), with some puncta abutting somatostatin (SST)-positive preBötC neurons (Fig. 1h and Extended Data Fig. 2). Approximately 90 preBötC neurons expressed *Nmbr*, the G-protein-coupled receptor specific for NMB (Fig. 1i, see later). Thus, *Nmb*-expressing RTN/pFRG neurons may directly modulate preBötC neurons.

NMB injection into preBötC induces sighing

To investigate the function of this NMB pathway, the peptide was micro-injected into preBötC of urethane-anaesthetized adult rats. Before injection, and after control (saline) injections (data not shown), airflow and diaphragmatic activity (DIA_{EMG}) showed the normal (eupneic) breathing pattern, with diaphragm activity bursts during inspiration (Fig. 2a, b and Extended Data Fig. 3a). Every minute or two, we observed a sigh ($44 \pm 10 \text{ h}^{-1}$, $n = 24$; unaffected by saline injection, Extended Data Fig. 3b), a biphasic double-sized breath coincident with a biphasic DIA_{EMG} event (Fig. 2b and Extended Data Fig. 3a, c–f). Amplitude

¹Department of Biochemistry and Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, California 94305, USA. ²Systems Neurobiology Laboratory, Department of Neurobiology, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California 90095, USA. [†]Present addresses: Department of Cell Biology and Anatomy, Chicago Medical School, Rosalind Franklin University of Medicine and Science, North Chicago, Illinois 60064, USA (K.K.); Department of Physiology, University of Alberta, Edmonton, Alberta T6G 2E1, Canada (S.P.).

*These authors contributed equally to this work.

§These authors jointly supervised this project.

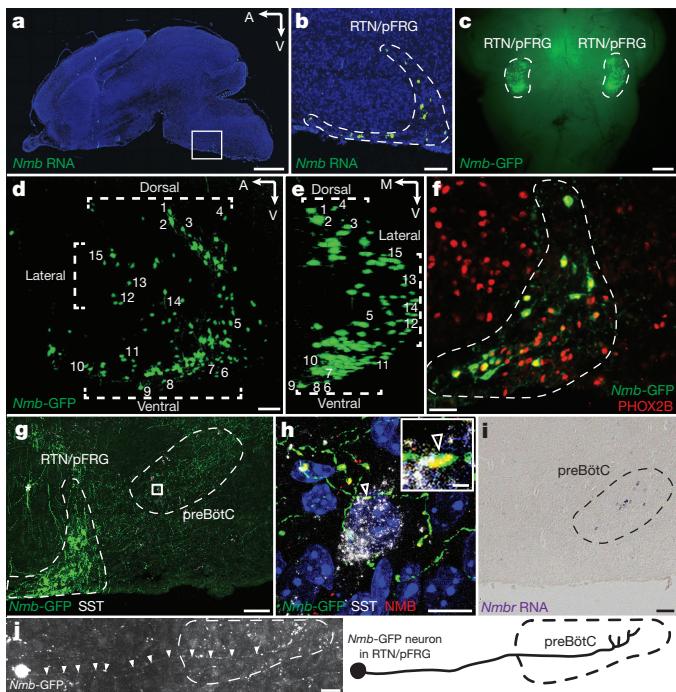


Figure 1 | NMB neuropeptide pathway neurons in breathing centre.

a, P0 mouse brain section probed for *Nmb* mRNA (green) with DAPI counterstain (nuclei, blue). Scale bar, 1 mm. **b**, Boxed region (a) showing specific expression in RTN/pFRG. Scale bar, 100 μ m. **c**, Whole mount P0 brainstem (ventral view) showing *Nmb*-GFP transgene expression (GFP, green) bilaterally in RTN/pFRG. Scale bar, 0.5 mm. **d, e**, Three-dimensional reconstruction (sagittal (d), coronal (e) projections) of CLARITY-cleared P14 *Nmb*-GFP brainstem. Note, RTN/pFRG expression ventral, dorsal, and lateral to facial nucleus. Numbers, representative neurons. A, anterior; V, ventral; M, medial. Scale bar, 100 μ m. **f**, P0 *Nmb*-GFP-expressing neurons (green) in RTN/pFRG (dashed) co-express RTN marker PHOX2B (red). Scale bar, 50 μ m. **g**, P7 *Nmb*-GFP-expressing neurons (green) project to preBötC (dashed). SST (somatostatin), preBötC marker (white). Asterisk indicates isolated GFP-labelled neuron in facial nucleus. Scale bar, 100 μ m. **h**, Boxed region (g) with NMB co-stain (z-stack projection; optical sections, Extended Data Fig. 2). Arrowhead, NMB puncta (red) in *Nmb*-GFP-expressing projection (green) abutting preBötC neuron (SST, white). Scale bars, 10 μ m (1 μ m, inset). **i**, P7 ventral medulla section probed for *Nmbr* mRNA (purple) showing preBötC expression. Scale bar, 100 μ m. **j**, Tiled image (left) and tracing (right) of *Nmb*-GFP neuron as in **g** projecting to preBötC. Scale bar, 30 μ m.

and timescale of the first component of a sigh was indistinguishable from eupneic breaths, like human sighs²⁶. Following bilateral NMB microinjection (100 nl, 3 μ M) into preBötC, sighing increased 6–17-fold ($n = 5$; Fig. 2a, c, d and Extended Data Fig. 4a–e). The effect peaked several minutes after injection and persisted for 10–15 min afterwards.

We also tested NMB on explanted preBötC brain slices of neonatal mice, where inspiratory activity is detected as rhythmic bursts of preBötC neurons and hypoglossal (cranial nerve XII) motoneuron output (Fig. 2e). Occasionally, a burst with two peaks ('doublet') was observed (Fig. 2e)¹¹, a proposed *in vitro* signature of a sigh (see Methods). Addition of 10 nM and 30 nM NMB increased doublet frequency by 1.7-fold ($P = 0.005$; $n = 7$) and twofold, respectively ($P = 0.003$; $n = 7$) (Fig. 2e, f). The overall frequency of bursts and doublets together was unchanged ($P = 0.2$; $n = 7$), implying NMB converts inspiratory bursts into sighs; indeed, in some preparations every inspiratory burst was converted to a doublet (Extended Data Fig. 5). We conclude that NMB acts directly on preBötC to increase sighing.

NMBR signalling maintains basal sighing

To determine if NMB signalling is required for sighing, we monitored breathing of awake, unrestrained *Nmbr*^{−/−} knockout mice. Wild-type

controls (C57BL/6) sighed 40 ± 11 h^{−1}, whereas *Nmbr*^{−/−} mutants sighed 29 ± 10 h^{−1} ($n = 4$; $P < 0.001$) (Fig. 2g). Sighing was also transiently reduced ~50% in anaesthetized rats by NMBR inhibition following bilateral preBötC injection of the antagonist BIM23042 (100 nl, 6 μ M) (Fig. 2h and Extended Data Fig. 6a–d and 7a). The antagonist effect was selective for sighing as it did not significantly alter respiratory rate (117 ± 14 vs 109 ± 6 breaths per minute with antagonist, $n = 4$, $P = 0.14$) or tidal volume (2.1 ± 0.1 vs 2.0 ± 0.3 ml, $n = 4$, $P = 0.46$), and similar selectivity was observed for the *Nmbr* mutant (respiratory rate 218 ± 22 vs 254 ± 22 in *Nmbr*^{−/−} mice, $n = 4$, $P = 0.06$; tidal volume 0.30 ± 0.03 vs 0.31 ± 0.04 ml, $n = 4$, $P = 0.88$). Thus NMBR signalling in preBötC maintains basal sighing.

A related neuropeptide pathway modulates sighing

Nmbr mutations and inhibition reduced but did not abolish sighing, suggesting involvement of other pathways. *Grp*, the only other bombesin-like neuropeptide gene in mammals²⁷, was expressed in several dozen cells in the dorsal RTN/pFRG in mouse (Fig. 3a) and rat (Extended Data Fig. 8e), plus scattered cells in nucleus tractus solitarius (NTS) and parabrachial nucleus (PBN), two other breathing circuit nuclei^{10,28} (Extended Data Fig. 8a–d). GRP-positive projections were traced from RTN/pFRG to preBötC, with some GRP puncta abutting SST-positive preBötC neurons (Extended Data Fig. 8f–i). GRP signals through GRPR, the receptor most similar to NMBR. *Grpr* mRNA was detected in ~160 mouse preBötC neurons (Fig. 3b and see later), suggesting GRP can also directly modulate preBötC function.

To determine if GRP regulates sighing, the neuropeptide (100 nl, 3 μ M) was injected bilaterally into preBötC of anaesthetized rats. Sighing increased 8–16-fold ($n = 5$; Fig. 3c and Extended Data Fig. 4f–j). GRP (3 nM) application to mouse preBötC brain slices also increased sighing, producing 1.7-fold more doublets ($P = 0.003$; $n = 9$; Fig. 3d). Thus, GRP can induce sighing through direct modulation of preBötC neurons, like NMB.

To determine if GRPR signalling is required for sighing, we monitored breathing in *Grpr*^{−/−} knockout mice. Their basal sigh rate (22 ± 9 per hour, $n = 4$) was half that of control wild-type mice (Fig. 3e), whereas eupneic breathing appeared normal (respiratory rate 218 ± 22 vs 210 ± 16 in *Grpr*^{−/−}, $n = 4$, $P = 0.57$; tidal volume 0.30 ± 0.03 vs 0.28 ± 0.01 ml, $n = 4$, $P = 0.23$). GRPR inhibition by bilateral preBötC injection of antagonist RC3095 (100 nl, 6 μ M) in anaesthetized rats also transiently decreased sighing by ~50% ($n = 4$), followed by rapid rebound and overshoot (Fig. 3f and Extended Data Fig. 6e–h). There was no significant change in other respiratory parameters (respiratory rate 117 ± 12 vs 111 ± 11 with antagonist, $n = 4$, $P = 0.34$; tidal volume 2.0 ± 0.2 vs 1.9 ± 0.1 ml, $n = 4$, $P = 0.11$). Thus, GRPR signalling in preBötC also maintains basal sighing.

Expression patterns, loss-of-function phenotypes and localized pharmacological manipulations of NMBR and GRPR signalling in preBötC suggest that NMB–NMBR and GRP–GRPR pathways can independently modulate sighing.

NMBR and GRPR are the critical pathways in sighing

To explore the relationship between NMB and GRP pathways, we compared expression patterns of the neuropeptides and receptors within mouse RTN/pFRG and preBötC. *Nmb* and *Grp* were detected in non-overlapping neuronal subpopulations, with *Nmb* neurons distributed throughout RTN/pFRG and *Grp* neurons restricted to the dorsal domain (Fig. 4a–d). In contrast, receptor expression patterns in preBötC overlapped (Fig. 4e–h), with 40 ± 16 neurons expressing *Nmbr*, 113 ± 45 expressing *Grpr*, and 49 ± 9 expressing both ($n = 3$).

To explore functional interactions, we injected both neuropeptides into preBötC of anaesthetized rats. Sigh rate increased 12–24-fold, which was similar or slightly beyond that of either neuropeptide alone (Fig. 4i and Extended Data Fig. 4k–o). When NMBR and GRPR pathways were simultaneously inhibited by bilateral injection of both antagonists, BIM23042 (100 nl, 6 μ M) and RC3095 (100 nl, 6 μ M), sighing was

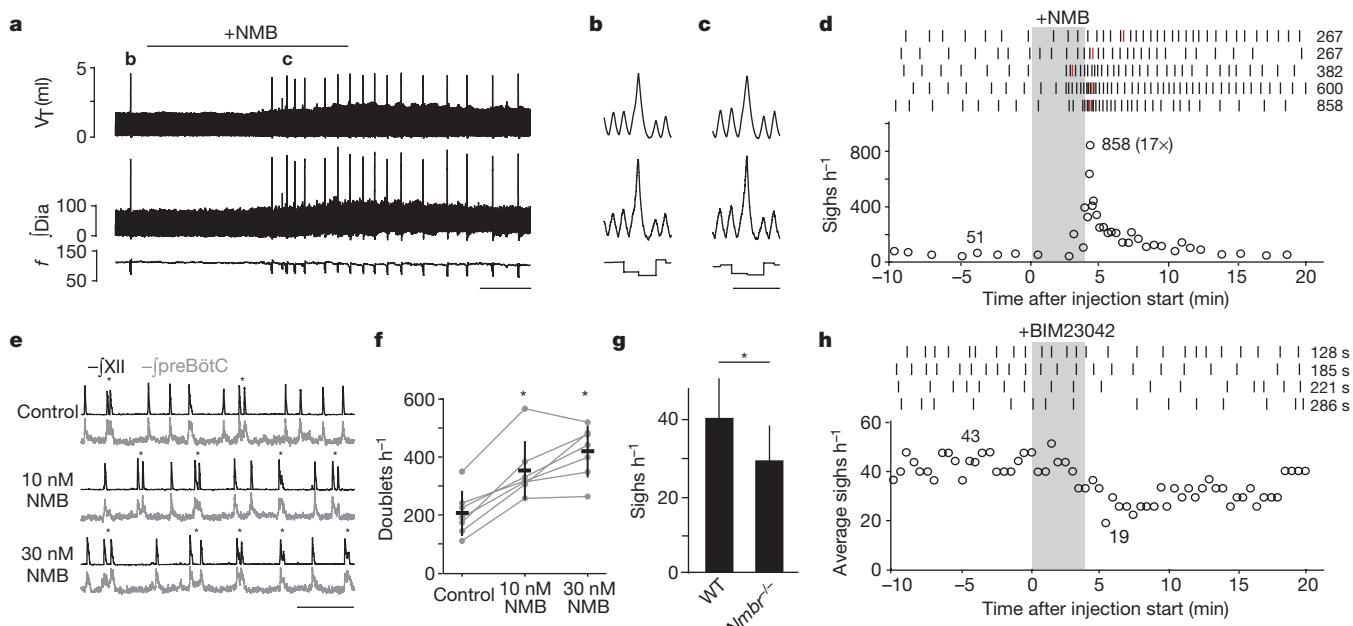


Figure 2 | NMB effect on breathing. **a–c**, Breathing activity of anaesthetized rat following bilateral NMB injection (100 nl, 3 μ M) into preBötC. Note increased sighing (spikes in tidal volume (V_T), integrated diaphragm activity (f/Dia)), but little change in respiratory rate (frequency, f). Bar, 1 min. **b, c**, Similar, stereotyped waveforms of spontaneous (**b**) and NMB-induced (**c**) sighs (from **a**; also Extended Data Fig. 3a, c–f). Bar, 2 s. **d**, Quantification of (**a**). Top: raster plots of sighs (tics) in five rats following NMB injection (grey); numbers, highest instantaneous sigh rate (red ticks). Bottom: instantaneous sigh rate of bottom raster plot; numbers, average instantaneous sigh rate before and maximum (and fold increase) after injection. **e**, Integrated hypoglossal nerve (f_{XII} ; black) and preBötC

neural activity ($f_{preBötC}$; grey) in preBötC slices containing indicated NMB concentrations. NMB increases doublets (*), a sigh signature in slices. Bar, 10 s. **f**, Quantification of (**e**) (data as mean \pm s.d. $n=7$; $^*P<0.05$ by paired *t*-test). **g**, Basal sigh rate in C57BL/6 wild-type (WT) and $Nmbr^{-/-}$ mice. $n=4$; data as mean \pm s.d.; $^*P<0.001$ by unpaired *t*-test. **h**, Effect on sighing in anaesthetized rats of bilateral preBötC injection (grey) of NMBR antagonist BIM23042 (100 nl, 6 μ M). Top: raster plots; numbers, longest inter-sigh intervals (s, seconds) following injection. Bottom: sliding average sigh rate (bin 4 min; slide 30 s); numbers, average rate before (left) and minimum binned rate after injection (right).

severely reduced or eliminated ($n=6$; Fig. 4j, Extended Data Fig. 6i–n). Thus, NMBR and GRPR pathways can independently modulate sighing, and together are required for basal sighing *in vivo*.

Effect of NMBR and GRPR neuron ablation

To determine if preBötC NMBR- and GRPR-expressing neurons function specifically in sigh control, we ablated them using bombesin (BBN), which binds both receptors²⁷, conjugated to saporin

(BBN–SAP), a ribosomal toxin that induces neuronal death when internalized⁸. Three days after bilateral BBN–SAP injection (200 nl, 6.15 ng per side) into preBötC of rats, sighing was reduced $\sim 80\%$, from $24 \pm 3 \text{ h}^{-1}$ before injection to $5 \pm 4 \text{ h}^{-1}$ three days after injection ($P=5 \times 10^{-6}$; $n=7$) (Fig. 5a). The effect was selective as other aspects of breathing and behaviour appeared normal. Five days after injection, sighing was almost completely ($\sim 95\%$) abolished, decreasing to $0.6 \pm 0.6 \text{ h}^{-1}$ ($P=10^{-8}$; $n=6$; Fig. 5a). Other aspects of breathing and

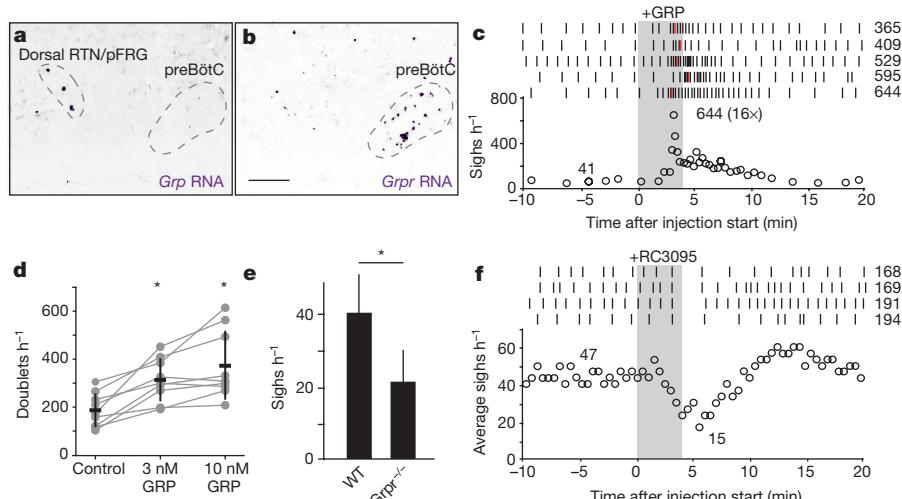


Figure 3 | GRP neuropeptide pathway expression and function in breathing. **a, b**, Sagittal ventral medulla sections of P7 mice probed for *Grp* (**a**) or *Grpr* (**b**) mRNA (purple). Scale bar, 200 μ m. **c**, Effect on sighing of bilateral preBötC injection of GRP (100 nl, 3 μ M), as in Fig. 2d. **d**, Effect of GRP on doublets (sighs) in preBötC slices, as in Fig. 2f. Data as

mean \pm s.d. $n=9$; $^*P<0.05$ by paired *t*-test. **e**, Basal sigh rate in C57BL/6 wild-type (WT) and $Grpr^{-/-}$ mice. $n=4$; data as mean \pm s.d.; $^*P<0.001$ by Mann–Whitney *U*-test. **f**, Effect on sighing of bilateral preBötC injection of GRP antagonist RC3095 (100 nl, 6 μ M), as in Fig. 2h.

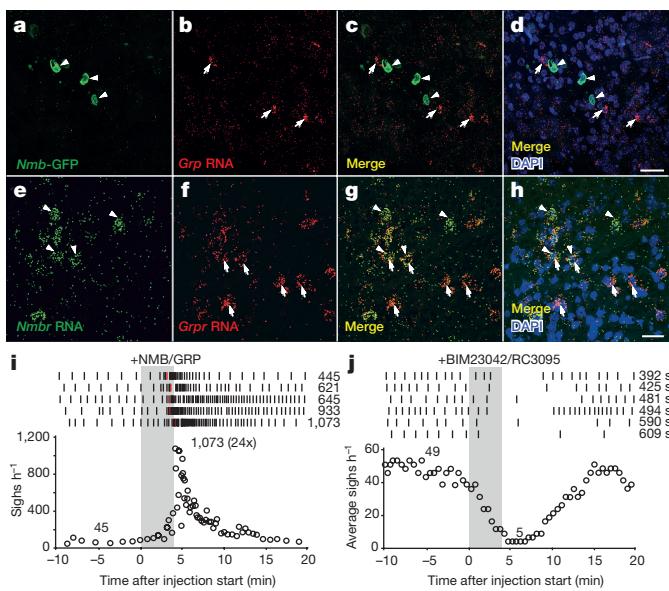


Figure 4 | Interactions between NMB and GRP pathways in sighing. a–d, RTN/pFRG section of P7 *Nmb*-GFP mouse immunostained for GFP (green, arrowheads) and probed for *Grp* mRNA (red, arrows). Note no expression overlap. Scale bar, 30 μ m. e–h, preBötC section of P28 mouse probed for *Nmbr* mRNA (green, arrowheads) and *Grpr* mRNA (red, arrows). Note partial expression overlap. Scale bar, 30 μ m. i, Effect on sighing of bilateral preBötC injection of both NMB (100 nl, 3 μ M) and GRP (100 nl, 3 μ M) as in Fig. 2d. j, Effect on sighing of bilateral preBötC injection (100 nl, 6 μ M) of both NMBR and GRPR antagonists (BIM23042, RC3095) as in Fig. 2h.

behaviour again appeared generally intact (Extended Data Fig. 9a, b). However, after 5 days we noted increasing episodes of apneas or disordered breathing, possibly a consequence of the loss of sighing (see Discussion). Ablation prevented sigh induction by exogenous BBN infusion into the cisterna magna, confirming preBötC *Nmbr*- and *Grpr*-expressing neurons were eliminated (Extended Data Fig. 9c). However, BBN infusion still triggered intense scratching and licking, demonstrating that the *Grpr* and *Nmbr* neurons outside the preBötC required for these behaviours²⁹ remained intact. We conclude that preBötC *Nmbr* and *Grpr*-expressing neurons have a critical and selective function in basal sighing.

NMBR and GRPR neurons are critical for induced sighs

To determine if the *Nmbr* and *Grpr*-expressing neurons are also important for physiologically-induced sighs, we examined BBN–SAP rats exposed to hypoxia (8% O₂). In control rats injected with unconjugated SAP, sighing increased from 24 ± 5 to 140 ± 14 h⁻¹ under hypoxia ($n = 3$, $P = 0.01$). In contrast, five days after BBN–SAP injection, sigh rate under hypoxia was 5.1 ± 7.9 h⁻¹ ($n = 6$, $P = 0.2$, hypoxia vs room air at day 5; Fig. 5b); three of these rats did not sigh in room air (21% O₂) and no sighs were triggered by hypoxia. In BBN–SAP rats, hypoxia increased respiratory rate from 150 ± 5 to 230 ± 19 breaths per min, demonstrating ventilatory response to hypoxia was intact. Thus, *Nmbr*-expressing and *Grpr*-expressing neurons are also critical for hypoxia-induced sighing, but not other respiratory responses to hypoxia.

Discussion

Our results show sighing is controlled by two largely parallel bombesin-like neuropeptide pathways, NMB and GRP, which mediate signalling between key medullary breathing control centres. Approximately 200 *Nmb*-expressing and ~ 30 *Grp*-expressing neurons in neighbouring domains of RTN/pFRG, a region implicated in integrating respiratory sensory cues and generation of active expiration^{23,28,30}, project to preBötC, the respiratory rhythm generator. A total of 7% (~ 200) of preBötC neurons express *Nmbr* (~ 40 neurons),

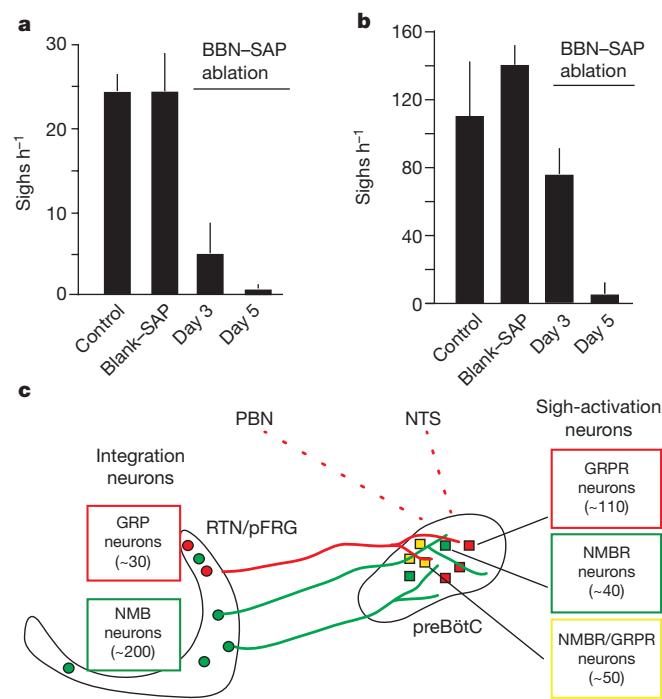


Figure 5 | Effect on sighing of ablating preBötC NMBR-expressing and GRPR-expressing neurons. a, b, Basal (a) and hypoxia-induced (b) sigh rates before (control) and 3 or 5 days after preBötC injections of bombesin–saporin (200 nl, 6.2 ng; BBN–SAP ablation) to ablate NMBR- and GRPR-expressing neurons, or 5 days after saporin alone (200 nl, 6.2 ng; Blank–SAP). Data as mean \pm s.d.; $P < 0.05$ (BBN–SAP ablation vs control rats at day 3 or 5 for both room air (a) and hypoxia (b)). Sample size $n = 6$ (control), 3 (Blank–SAP), 7 (day 3), 6 (day 5). c, Model of peptidergic sigh control circuit. NMB- and GRP-expressing neurons in RTN/pFRG (and perhaps GRP-expressing neurons in NTS and PBN) receive physiological and perhaps emotional input from other brain regions, stimulating neuropeptide secretion. This activates receptor-expressing preBötC neurons expressing their receptors, which transform the normal preBötC rhythm to sighs. (Because neuropeptides induce sighs separated by normal breaths (Fig. 2a), there must be some refractory mechanism in or downstream of receptor-expressing neurons that temporarily prevents a second sigh.)

Grpr (~ 110) or both receptors (~ 50), activation of which increased sighing 6-fold to 24-fold, whereas sighing was effectively abolished by inhibition or deletion of the receptors, or ablation of the receptor-expressing neurons.

We propose the above neurons, perhaps with *Grp*-expressing NTS and PBN neurons, comprise the core of a peptidergic sigh control circuit, with the neuropeptide-expressing neurons integrating inputs from sites monitoring physiological and perhaps emotional state (Fig. 5c). Excitation of these neurons and secretion of either neuropeptide activates the cognate receptor-expressing preBötC neurons, which initiate sighs by altering activity of other preBötC neurons to convert normal breaths to sighs. This might occur by a burst of NMB and/or GRP secretion triggering a second inspiratory signal in the preBötC during or immediately after the first, resulting in a single, double-sized breath. Alternatively, NMB and GRP secretion might be more gradually modulated, causing concentration-dependent bursts in activity of receptor-expressing neurons or a shift in preBötC properties towards states favouring more frequent doublet bursts.

A priority will be to elucidate the full circuit and properties of the constituent neurons, including its integration with other peptides and neurotransmitters that influence sighing^{15,16} and with normal breathing and other behaviours. A curious aspect of the circuit already apparent is the central role of two partially overlapping and closely related neuropeptide pathways. Do NMB- and GRP-expressing neurons receive different inputs and have distinct sensing functions, and do the three

sets of receptor-expressing neurons (NMBR, GRPR, NMBR + GRPR) converge on the same preBötC neurons to effect a sigh, or signal to different preBötC neurons producing distinct types of sighs?

A striking aspect of our results is the selectivity of the circuit for sighing. Inhibition of the pathways, and even ablation of receptor-expressing neurons, had little effect on other aspects of breathing, at least in the short term. This can now be exploited to test the classical hypothesis for the physiological function of sighing—re-expansion of alveoli that collapse during breathing and maintenance of lung integrity^{2–4}, and to investigate psychological benefits. Identification of the key neuropeptide pathways suggests pharmacologic approaches for controlling excessive sighing and inducing sighs in patients that cannot breathe deeply on their own. Dozens of molecularly distinct preBötC neuronal types have been identified recently (K.Y. and M.A.K., in preparation); perhaps they serve similarly specific roles in other respiratory-related behaviours like yawning, sniffing, crying, and laughing.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 28 May; accepted 30 December 2015.

Published online 8 February 2016.

1. Haldane, J. S., Meakins, J. C. & Priestley, J. G. The effects of shallow breathing. *J. Physiol. (Lond.)* **52**, 433–453 (1919).
2. McCutcheon, F. H. Atmospheric respiration and the complex cycles in mammalian breathing mechanisms. *J. Cell. Physiol.* **41**, 291–303 (1953).
3. Knowlton, G. C. & Larabee, M. G. A unitary analysis of pulmonary volume receptors. *Am. J. Physiol.* **147**, 100–114 (1946).
4. Reynolds, L. B. Characteristics of an inspiration-augmenting reflex in anesthetized cats. *J. Appl. Physiol.* **17**, 683–688 (1962).
5. Bartlett, D. Origin and regulation of spontaneous deep breaths. *Respir. Physiol.* **12**, 230–238 (1971).
6. Maytum, C. K. Sighing dyspnea: A clinical syndrome. *J. Allergy* **10**, 50–55 (1938).
7. Smith, J. C., Ellenberger, H. H., Ballanyi, K., Richter, D. W. & Feldman, J. L. Pre-Bötziinger complex: a brainstem region that may generate respiratory rhythm in mammals. *Science* **254**, 726–729 (1991).
8. Gray, P. A., Janczewski, W. A., Mellen, N., McCrimmon, D. R. & Feldman, J. L. Normal breathing requires preBötziinger complex neurokinin-1 receptor-expressing neurons. *Nature Neurosci.* **4**, 927–930 (2001).
9. Tan, W. et al. Silencing preBötziinger complex somatostatin-expressing neurons induces persistent apnea in awake rat. *Nature Neurosci.* **11**, 538–540 (2008).
10. Feldman, J. L., Del Negro, C. A. & Gray, P. A. Understanding the rhythm of breathing: so near, yet so far. *Annu. Rev. Physiol.* **75**, 423–452 (2013).
11. Kam, K., Worrell, J. W., Janczewski, W. A., Cui, Y. & Feldman, J. L. Distinct inspiratory rhythm and pattern generating mechanisms in the preBötziinger complex. *J. Neurosci.* **33**, 9235–9245 (2013).
12. Lieske, S. P., Thoby-Brisson, M., Telkamp, P. & Ramirez, J. M. Reconfiguration of the neural network controlling multiple breathing patterns: eupnea, sighs and gasps. *Nature Neurosci.* **3**, 600–607 (2000).
13. Ruangkittisakul, A. et al. Generation of eupnea and sighs by a spatiochemically organized inspiratory network. *J. Neurosci.* **28**, 2447–2458 (2008).
14. Caughey, J. L. Jr. Analysis of breathing patterns. *Am. Rev. Tuberc.* **48**, 382 (1943).
15. Niewoehner, D. E., Levine, A. S. & Morley, J. E. Central effects of neuropeptides on ventilation in the rat. *Peptides* **4**, 277–281 (1983).

16. Ramirez, J. M. The integrative role of the sigh in psychology, physiology, pathology, and neurobiology. *Prog. Brain Res.* **209**, 91–129 (2014).
17. Janczewski, W. A., Pagliardini, S., Cui, Y. & Feldman, J. L. Sighing after vagotomy. *Proc. Society of Neuroscience Conference* (New Orleans, 2012).
18. Smith, J. C., Morrison, D. E., Ellenberger, H. H., Otto, M. R. & Feldman, J. L. Brainstem projections to the major respiratory neuron populations in the medulla of the cat. *J. Comp. Neurol.* **281**, 69–96 (1989).
19. Onimaru, H. & Homma, I. A novel functional neuron group for respiratory rhythm generation in the ventral medulla. *J. Neurosci.* **23**, 1478–1486 (2003).
20. Diez-Roux, G. et al. A high resolution atlas of the transcriptome in the mouse embryo. *PLoS Biol.* **9**, e1000582 (2011).
21. Tomer, R. et al. Advanced CLARITY for rapid and high-resolution imaging of intact tissues. *Nature Protocols* **9**, 1682–1697 (2014).
22. Mulkey, D. K. et al. Respiratory control by ventral surface chemoreceptor neurons in rats. *Nature Neurosci.* **7**, 1360–1369 (2004).
23. Stornetta, R. L. et al. Expression of Phox2b by brainstem neurons involved in chemosensory integration in the adult rat. *J. Neurosci.* **26**, 10305–10314 (2006).
24. Guyenet, P. G. & Bayliss, D. A. Neural control of breathing and CO₂ homeostasis. *Neuron* **87**, 946–961 (2015).
25. Lazarenko, R. M. et al. Acid sensitivity and ultrastructure of the retrotrapezoid nucleus in Phox2b-EGFP transgenic mice. *J. Comp. Neurol.* **517**, 69–86 (2009).
26. Bendixen, H. H., Smith, G. M. & Mead, J. Pattern of ventilation in young adults. *J. Appl. Physiol.* **19**, 195–198 (1964).
27. Jensen, R. T. et al. International Union of Pharmacology. LXVIII. Mammalian bombesin receptors: nomenclature, distribution, pharmacology, signaling, and functions in normal and disease states. *Pharmacol. Rev.* **60**, 1–42 (2008).
28. Guyenet, P. G., Stornetta, R. L. & Bayliss, D. A. Central respiratory chemoreception. *J. Comp. Neurol.* **518**, 3883–3906 (2010).
29. Lee, H., Naughton, N. N., Woods, J. H. & Ko, M. C. Characterization of scratching responses in rats following centrally administered morphine or bombesin. *Behav. Pharmacol.* **14**, 501–508 (2003).
30. Pagliardini, S. et al. Active expiration induced by excitation of ventral medulla in adult anesthetized rats. *J. Neurosci.* **31**, 2895–2905 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank M. Sunday for providing the *Nmbr*^{−/−} and *Grpr*^{−/−} mice, Y. Zhang for providing rat tissues, and K. Wada and E. Wada for plasmid constructs for *in situ* hybridization probes. We also thank members of the Krasnow and Feldman laboratories for comments. This work was supported by the Howard Hughes Medical Institute (M.A.K.), NIH grants HL70029, HL40959 and NS72211 (J.L.F.), a Walter V. and Idun Berry postdoctoral fellowship (P.L.), the NIH Medical Scientist Training Program (K.Y.), and CIHR and AIHS postdoctoral fellowships (S.P.). M.A.K. is an investigator of the Howard Hughes Medical Institute.

Author Contributions W.A.J. and S.P. performed experiments showing the effects on sighing of bombesin injection into the preBötC and ablation of receptor-expressing neurons with bombesin-saporin. K.Y. performed the screen that discovered *Nmb* expression in the respiratory centres. P.L. performed experiments identifying and characterizing expression of *Nmb*, *Grp* and their receptors. W.A.J., P.L., and K.Y. performed genetic and pharmacology experiments on *Nmb* and *Grp* pathways. K.K. performed *in vitro* slice experiments. W.A.J., K.K., P.L., S.P., and K.Y. analysed data. J.L.F., W.A.J., K.K., M.A.K., P.L., and K.Y. conceived experiments, interpreted data and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.A.K. (krasnow@stanford.edu) or J.L.F. (feldman@g.ucla.edu).

METHODS

Animals. All procedures were carried out in accordance with animal care standards in National Institutes of Health (NIH) guidelines, and approved by the University of California, Los Angeles Animal Research Committee, or Stanford Institutional Animal Care and Use Committee. All mouse strains used were in the C57BL/6 genetic background. *Nmb*-GFP BAC transgenic mice (Tg(Nmb-EGFP)IT50Gsat/Mmuucd), carrying EGFP coding sequence inserted upstream of the *Nmb* start codon in the BAC, were from the Mutant Mouse Regional Resource Centers (catalogue number 030425-UCD, <https://www.mmrrc.org>). *Nmbr*^{-/-} and *Grpr*^{-/-} null mutant mice, in which exon 2 of the endogenous gene was replaced by a neomycin-resistance cassette using homologous recombination, have been described^{31,32}. Male Sprague Dawley rats were from Charles River.

In situ hybridization, immunostaining and reporter expression. For *in situ* hybridization, mouse brains were harvested and fixed overnight in 4% paraformaldehyde in phosphate buffered saline (PBS), cryopreserved in 30% sucrose and embedded in optical cutting temperature compound (OCT). Transverse sections were cut at 16–20 µm and stored at –80 °C until use. Sections were post-fixed in 4% paraformaldehyde before treatment with hydrochloric acid, proteinase K, and then triethanolamine/acetic anhydride. Hybridization was carried out with *in vitro* transcribed and digoxigenin-labelled riboprobes at 58 °C overnight. Signal was detected using alkaline phosphatase-coupled anti-digoxigenin (DIG) primary antibody (Roche) and nitro blue tetrazolium chloride and bromochloroindolyl phosphate (NBT/BCIP) Reagent Kit (Roche) or using Horse Radish Peroxidase-coupled anti-DIG primary antibody (Roche) and Tyramide Signal Amplification plus Fluorescent Substrate Kit (PerkinElmer).

For double fluorescent *in situ* hybridization, tissue was harvested, embedded in OCT and then sectioned. Sections were fixed in 4% paraformaldehyde, dehydrated and treated with pretreatment reagent (Advanced Cell Diagnostics). Double fluorescent *in situ* assay was then performed using proprietary RNAscope technology (Advanced Cell Diagnostics) with cyanine 3 labelled *Nmbr* probes and fluorescein isothiocyanate labelled *Grpr* probes.

For *Nmb*-GFP reporter expression analysis, the brains of *Nmb*-GFP mice were harvested and fixed overnight in 4% paraformaldehyde and then cryopreserved at 4 °C in 30% sucrose overnight. Tissue was embedded in OCT and sectioned at 10–40 µm. Tissue sections were rinsed with PBT (PBS + 0.1% Tween), blocked with 3% bovine serum (BSA) in PBT for 1 h, and incubated with primary antibody overnight at 4 °C. Sections were rinsed in PBT and incubated for 1 h at room temperature with species-specific secondary antibodies. Primary antibodies were: chicken anti-GFP (Abcam 13970; used at 1:1,000 dilution), goat anti-PHOX2B (Santa Cruz, sc-13224; 1:200 dilution), rabbit anti-NMB (Sigma-Aldrich, SAB1301059; 1:100 dilution), rabbit anti-GRP (Immunostar 20073; 1:4,000 dilution), and rat anti-SST (Millipore, MAB354; 1:50 dilution). Secondary antibodies included donkey anti-chicken (Jackson Immuno Research; 1:400 dilution), donkey anti-rat (Jackson Immuno Research; 1:400 dilution), donkey anti-rabbit (Jackson Immuno Research; 1:400 dilution) and donkey anti-goat (Invitrogen; 1:500 dilution).

For *Nmb*-GFP expression analysis in samples prepared by CLARITY²¹, *Nmb*-GFP mice were perfused with PBS and formaldehyde-acrylamide hydrogel, and brain tissue was harvested and incubated in hydrogel monomer solution at 4 °C for 3 days. Tissue was then embedded in polymerized hydrogel by raising the temperature to 37 °C for 3 h. Blocks of 1 mm thickness were cut and washed in 4% sodium dodecyl sulphate (SDS) in sodium borate buffer at 37 °C for 2 to 3 weeks. Samples were washed with PBST for 2 days and incubated in FocusClear (CelExplorer), and GFP fluorescence was imaged on a Zeiss LSM780 confocal microscope.

Sigh monitoring and analysis. For awake animals, individual animals were placed in a whole body plethysmography chamber (Buxco) at room temperature (22 °C) in 21% O₂ (for normoxia) or 8% O₂ (for hypoxic challenge) balanced with N₂. Sighs were identified in plethysmography traces by the characteristic biphasic ramp, the augmented flow in the second phase of the inspiratory effort and the prolongation of expiratory time following the event. Sighs were also confirmed by visual monitoring of breathing behaviour. Given the high amplitude and distinctive waveform of sighs relative to standard eupneic breaths, sighs were unambiguously identified by both visual and computer-assisted scoring; no difference was detected in direct comparisons between methods or observers, so visual scoring was used. Female 8-week-old mice (*Nmbr*^{-/-}, *Grpr*^{-/-}, or C57BL/6 as wild-type control) were acclimated to the chamber for 10 min, and then the first fifteen recorded sighs were used to calculate the sigh rate. Similar sigh rates were observed for each animal when assayed on different days. Rats were acclimated to the chamber for ~1 h, and then baseline sigh rate and respiratory frequency were calculated for the next 2 h. For hypoxic (8% O₂) challenge, analysis was continued for 30 min under the hypoxic condition.

For anaesthetized rats, the trachea was cannulated and connected to a pneumotachograph (GM Instruments) to record airflow. A flow calibration was performed after every experiment along with a calculation of tidal volume (V_T) by digital integration. To monitor diaphragm activity, wire electrodes (Cooner Wire) were implanted into the diaphragm and electromyogram (EMG) signal sampled at 2 kHz (Powerlab 16SP; AD Instruments). Signal was rectified and digitally integrated (time constant of 0.1) to obtain a moving average using LabChart Pro 8 (AD Instruments) and Igor Pro 6 (Data Matrix) software. Sighs were identified in the airflow measurements as above and validated by double peaks in the EMG recordings.

preBötC injection of NMB and GRP agonists and antagonists. Male Sprague Dawley rats (*n* = 24) weighing 320–470 g were anaesthetized with urethane (1.5 g per kg), isoflurane (0.3–0.7 vol%), and ketamine (20 mg per kg per hour) and injected i.p. with atropine (0.3 mg per kg), then placed in a supine position in a stereotaxic instrument (David Kopf Instruments). A tracheostomy tube was placed in the trachea through the larynx, and the basal aspect of the occipital bone was removed to expose the ventral medulla. Injections were placed 750 µm caudal from the most rostral root of the hypoglossal nerve (RXII), 2 mm lateral to the midline, and 700 µm dorsal to the ventral medullary surface. Small corrections were made to avoid puncturing blood vessels on the medulla. Microinjection was done using a series of pressure pulses (Picospirizer; Parker-Hannifin) applied to the open end of micropipettes, with air pressure set so that each pulse ejected ~5 nl and a total volume of 0.1 µl was injected on each side. The concentration of injected neuropeptides and antagonists were: NMB, 3 µM; GRP, 3 µM; NMB and GRP, 3 µM each; BIM23042 (ref. 33), 6 µM; RC3095 (ref. 34), 6 µM; BIM23042 and RC3095, 6 µM each. To verify the accuracy of the injections, fluorescent polystyrene beads (0.2 µm FluoSpheres (Invitrogen; catalogue #F8811, #F8763 or #F8807); 2–5% vol) were added to the injected solutions, and following injection and physiological measurements the location of the fluorescent beads was visualized in wet tissue double stained for reelin and choline acetyltransferase (ChAT) to identify the preBötC.

Ablation of NMBR and GRPR-expressing preBötC neurons. Bilateral preBötC injections of saporin conjugated with either bombesin (BBN-SAP; Advanced Targeting System; 200 nl, 6.2 ng) or a non-targeted peptide (blank SAP; Advanced Targeting System; 200 nl, 6.2 ng) were performed in 300–350 g rats under anaesthesia (ketamine (90 mg per kg), and xylazine (10 mg per kg), administered i.p) using standard aseptic procedures. Rats were positioned on a stereotaxic frame with bregma 5 mm below lambda. The occipital bone was exposed and a small window was opened to perform BBN-SAP injections with a 40 µm diameter tip glass pipette inserted into the preBötC. Coordinates were (in mm): 0.9 rostral, 2.0 lateral, and 2.8 ventral to the obex. The electrode was left in place for 5 min after injection to minimize backflow of solution up the electrode track. After injection, a fine polyethylene cannula was implanted and cemented to the occipital bone to deliver BBN into the fourth ventricle. Neck muscles and skin were sutured back at the end of the surgery and rats were allowed to recover with pain medication, food and water ad libitum. Blank-SAP and BBN-SAP treated rats were tested for hypoxia (8% O₂ balanced with nitrogen, 30 min challenge) five days after surgery. Blank-SAP and BBN-SAP treated rats were also tested for response to BBN infusion in the cistern magna six days after surgery. The cannula implanted in the fourth ventricle was connected to a fine polyethylene tubing under isoflurane anaesthesia, and after recovery and placement of rats in a plethysmographic chamber, 10 µg of BBN diluted in 20 µl sterile saline was delivered followed by a 20 µl saline washout. Sigh rate and respiratory rate were calculated for 30 min following infusion and compared to pre-infusion values.

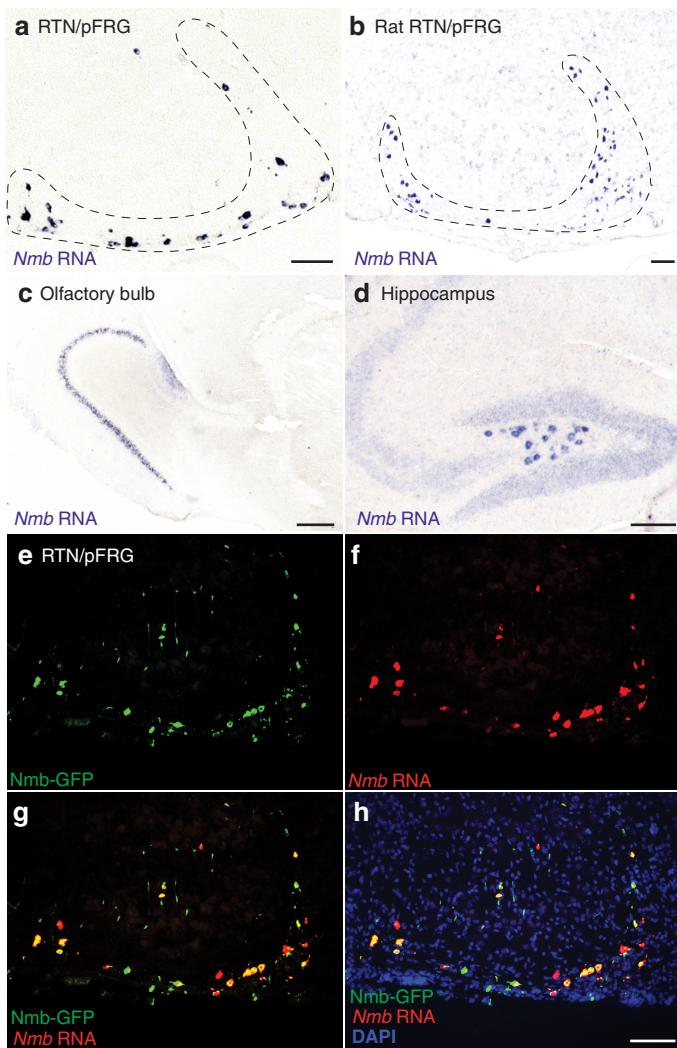
In vitro slice preparation, recording, and analysis. Rhythmic 550-µm-thick transverse medullary slices containing the preBötC and XII nerve from neonatal C57BL/6 mice (P0–5) were prepared as described previously¹¹. The medullary slice was cut in artificial cerebrospinal fluid (ACSF) containing (in mM): 124 NaCl, 3 KCl, 1.5 CaCl₂, 1 MgSO₄, 25 NaHCO₃, 0.5 NaH₂PO₄, and 30 d-glucose, equilibrated with 95% O₂ and 5% CO₂ (4 °C, pH = 7.4). For recording, extracellular K⁺ was raised to 9 mM to replace excitatory afferent drive lost in the cutting process. Slices were perfused at 27 °C and 4 ml min⁻¹ and allowed to equilibrate for 30 min. Respiratory activity reflecting suprathreshold action potential (AP) firing from populations of neurons was recorded as XII bursts from either XII nerve roots and as population activity directly from the preBötC using suction electrodes and a MultiClamp 700A or 700B (Molecular Devices, Sunnyvale, CA, USA), filtered at 2–4 kHz and digitized at 10 kHz. Digitized data were analysed off-line using custom procedures written for IgorPro (Wavemetrics, Portland, OR, USA). Activity was full-wave rectified and digitally integrated with a Paynter filter with a time constant of 20 ms with either custom built electronics or using custom procedures in MATLAB (Mathworks, Natick, MA, USA).

Burst detection and analysis of respiratory-related activity recorded in full-wave rectified XII output or preBötC population recordings were performed using custom software written in IgorPro. Burst parameters were normalized to the mode of the data in the baseline condition. Although there are several proposed definitions of sighs in slice preparations^{12,13}, here we used 'doublets' (double-peaked bursts) as the *in vitro* signature of sighs because, in our preparations, doublets detected both in preBötC neural population activity measurements and cranial nerve XII output recordings shared the increased inspiratory and expiratory duration of sighs¹¹. Furthermore, as demonstrated here, doublet rate increased following application of NMB or GRP to preBötC slice preparations (Figs 2 and 3 and Extended Data Fig. 5), as did sigh rate *in vivo* following preBötC injection of the same neuropeptides (Figs 2 and 3). The frequency and waveform of doublets in slice preparations does not closely match those of sighs in intact animals, presumably due to the absence *in vitro* of important inputs modulating burst shape; indeed, the doublets more closely resemble sighs in vagotomized animals, where they appear as equal amplitude double-peaked breaths^{11,17}. We scored a burst as a doublet if the burst displayed a second peak that reached 20% or more of the amplitude of the first burst, and this second peak occurred after more than twice the time from start to peak or if the burst had a duration longer than eight times the time from start to peak. All doublets were verified by visual inspection to exclude multipeaked bursts and two bursts that were too far apart. Measured doublet intervals were converted to a calculated per hour doublet rate.

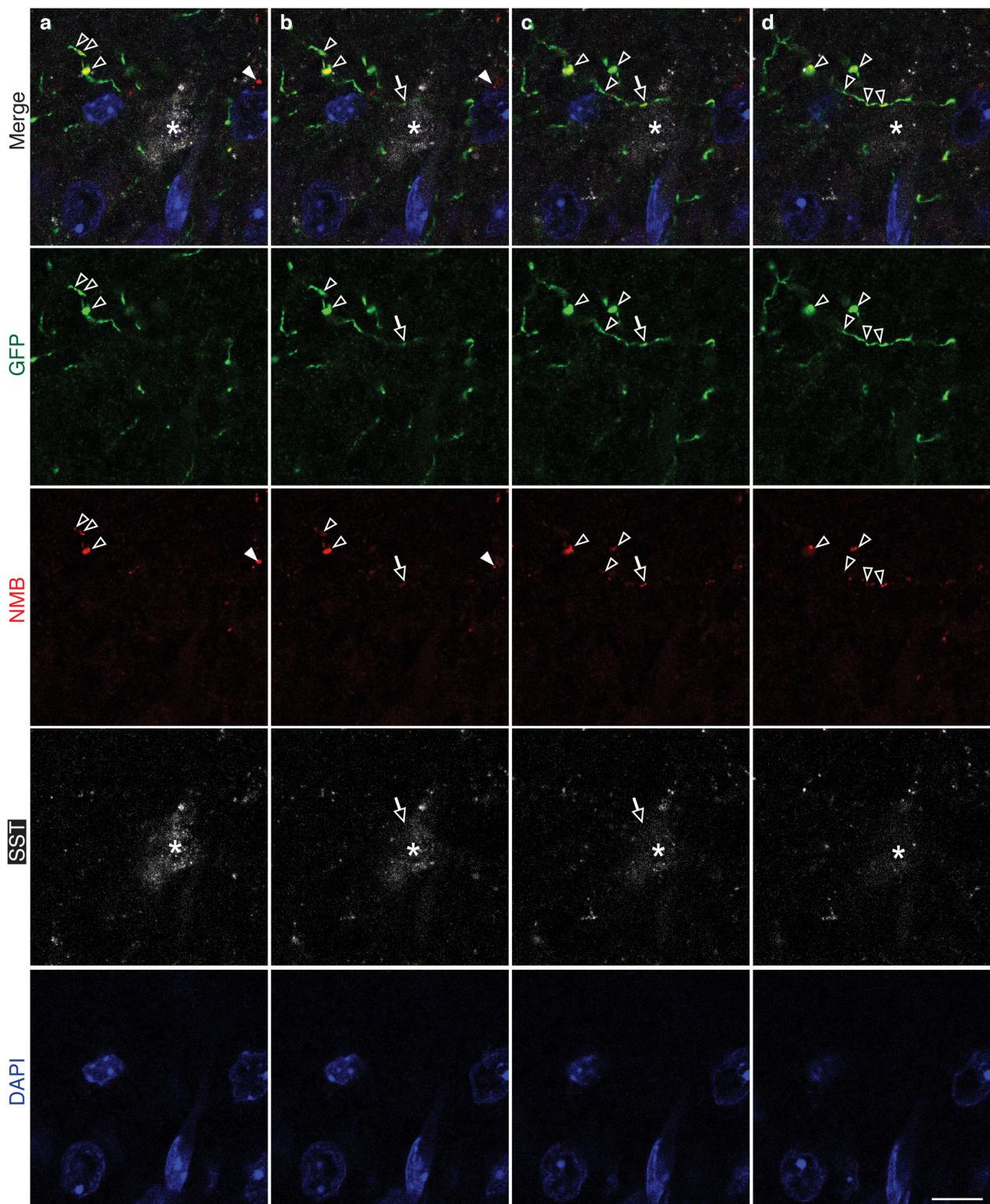
Statistics. Data are represented as mean \pm standard deviation (s.d.). Statistical significance was uniformly set at a minimum of $P < 0.05$. For comparisons of two groups, the assumption for normal distribution was determined by the

Shapiro–Wilk test with the critical W value set at 5% significance level. The t -tests were conducted, with the exception in Fig. 3e, in which a Mann–Whitney U -test was used. For statistical comparisons of more than two groups, an ANOVA was first performed. In most cases, a two-way repeated measures ANOVA was used for comparisons of various parameters in different conditions and for making comparisons across different events. If the null hypothesis (equal means) was rejected, post-hoc paired t -tests were then used for pairwise comparisons of interest. Individual P values are reported, but Holm–Bonferroni analysis for multiple comparisons was conducted to correct for interactions between the multiple groups. Histograms were normalized by the total sample size to generate plots of the relative frequency of each value where the y value of each bin represents the fraction of the total number of samples for that experiment. Randomization and blinding were not used. No statistical method was used to predetermine sample size.

31. Ohki-Hamazaki, H. *et al.* Functional properties of two bombesin-like peptide receptors revealed by the analysis of mice lacking neuromedin B receptor. *J. Neurosci.* **19**, 948–954 (1999).
32. Wada, E. *et al.* Generation and characterization of mice lacking gastrin-releasing peptide receptor. *Biochem. Biophys. Res. Commun.* **239**, 28–33 (1997).
33. Orbuch, M. *et al.* Discovery of a novel class of neuromedin B receptor antagonists, substituted somatostatin analogues. *Mol. Pharmacol.* **44**, 841–850 (1993).
34. Qin, Y. *et al.* Inhibitory effect of bombesin receptor antagonist RC-3095 on the growth of human pancreatic cancer cells *in vivo* and *in vitro*. *Cancer Res.* **54**, 1035–1041 (1994).

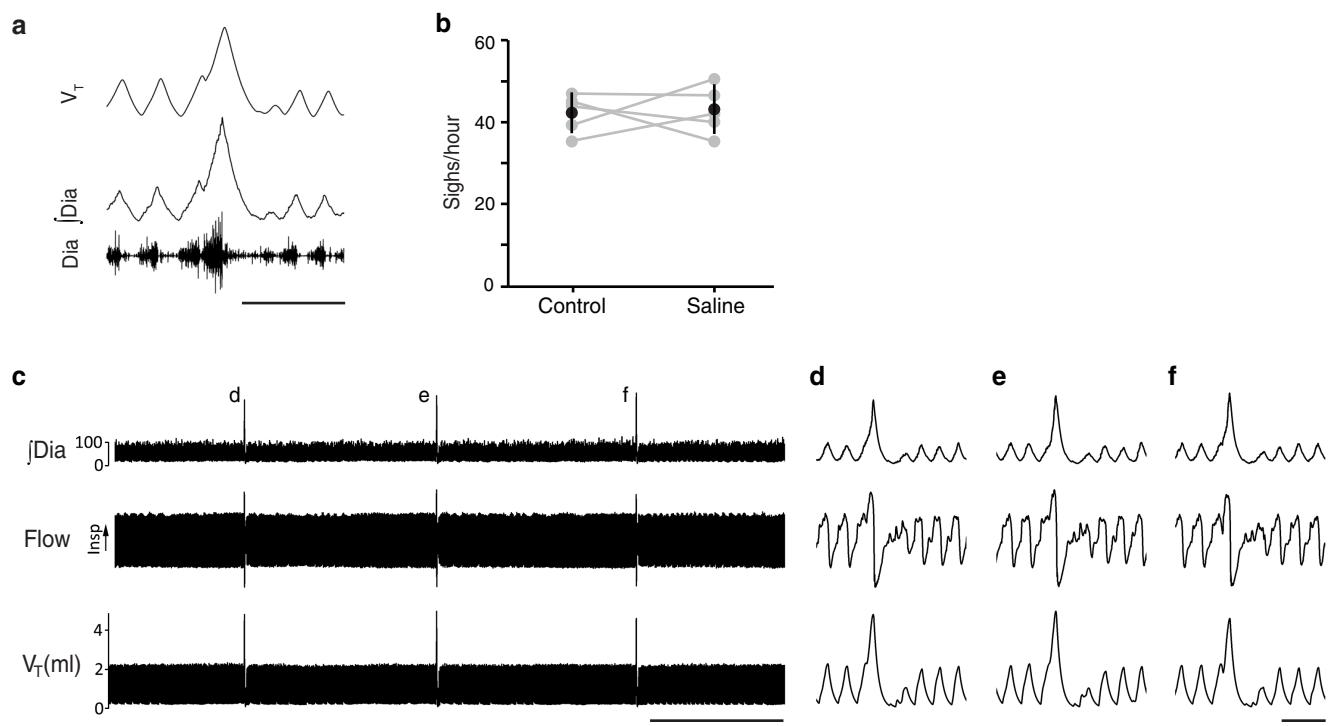


Extended Data Figure 1 | Expression of *Nmb* in rodent brain.
a, b, Sagittal sections of P7 mouse (a) and P7 rat (b) brain showing RTN/pFRG region probed for *Nmb* mRNA expression (purple) by *in situ* hybridization as in Fig. 1. Scale bars, 100 μ m. **c, d**, *Nmb* expression as in a showing regions outside ventrolateral medulla. *Nmb* is expressed in mouse olfactory bulb (c) and hippocampus (d). Scale bars, 200 μ m (c) and 100 μ m (d). **e-h**, Section through RTN/pFRG brain region of P0 transgenic *Nmb*-GFP mouse immunostained for GFP (green) and probed for *Nmb* mRNA (red) by *in situ* hybridization. Blue, DAPI nuclear stain. *Nmb*-GFP and *Nmb* mRNA are mainly co-expressed in the same cells. Scale bar, 100 μ m.



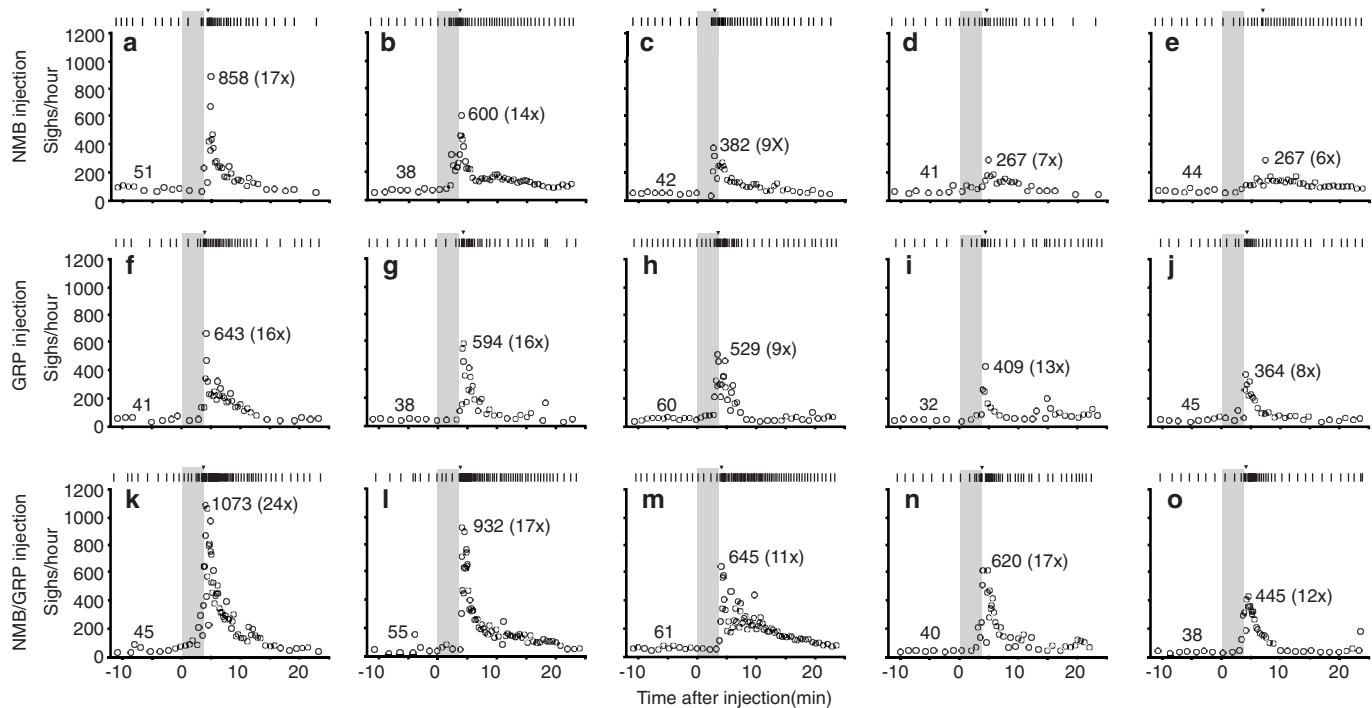
Extended Data Figure 2 | Serial confocal preBötC sections showing *Nmb*-GFP projections contain puncta of NMB. a-d, Serial confocal optical sections (0.6 μ m apart) through preBötC brain region of *Nmb*-GFP mouse immunostained for GFP (green), NMB (red), preBötC marker SST (white), and DAPI (blue) as in Fig. 2h. Note the GFP-positive projection with a puncta of NMB (yellow, open arrows in b, c) directly abutting an

SST positive neuron (asterisk). Most NMB puncta (open arrowheads) were detected within GFP-positive projections as expected, and only a small fraction of NMB puncta (closed arrowhead) were detected outside them; NMB outside *Nmb*-GFP projections could be secreted protein or the rare *Nmb*-expressing cells that do not co-express the *Nmb*-GFP transgene (see Extended Data Fig. 1e-h). Scale bar, 20 μ m.



Extended Data Figure 3 | Sighing after surgery and bilateral injection of saline into preBötC. **a**, Example of a sigh in breathing activity trace of a urethane anaesthetized rat after surgery as in Fig. 2a–c. V_T , tidal volume; $\int Dia$, integrated diaphragm activity; Dia, raw diaphragm activity trace. **b**, Sigh rate before (control) and after (saline) bilateral saline injection

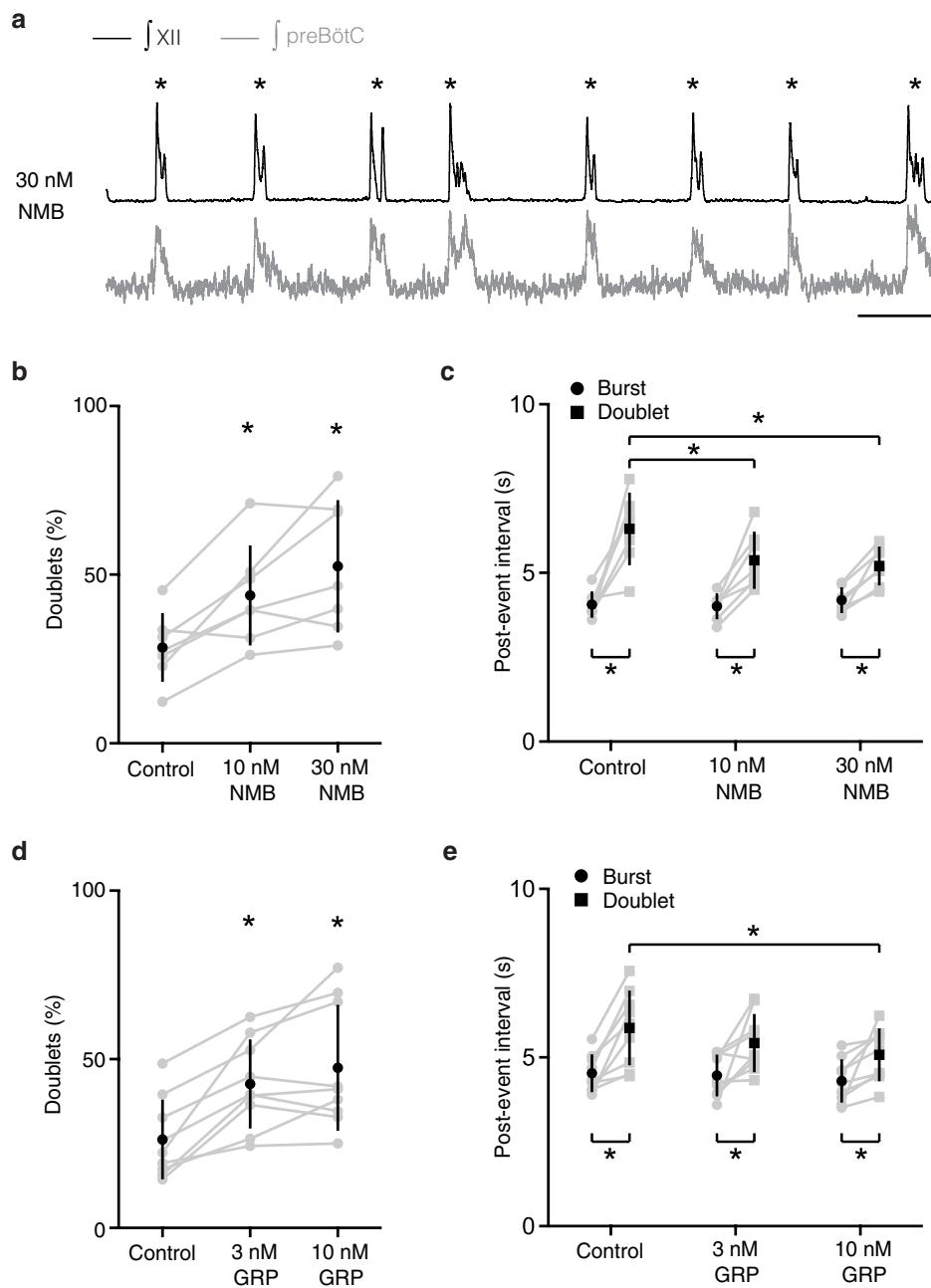
into preBötC. There is no effect of saline injection (data are mean \pm s.d., $n = 5$, $P = 0.83$ by paired t -test). **c–f**, Breathing activity trace as in **a** (but also showing airflow). Note stereotyped waveform of sighs (**d–f**). Bars, 1 min (**c**), 1 s (**d–f**).



Extended Data Figure 4 | Effects on sighing in individual rats following bilateral injection into preBötC of NMB, GRP and both NMB/GRP.

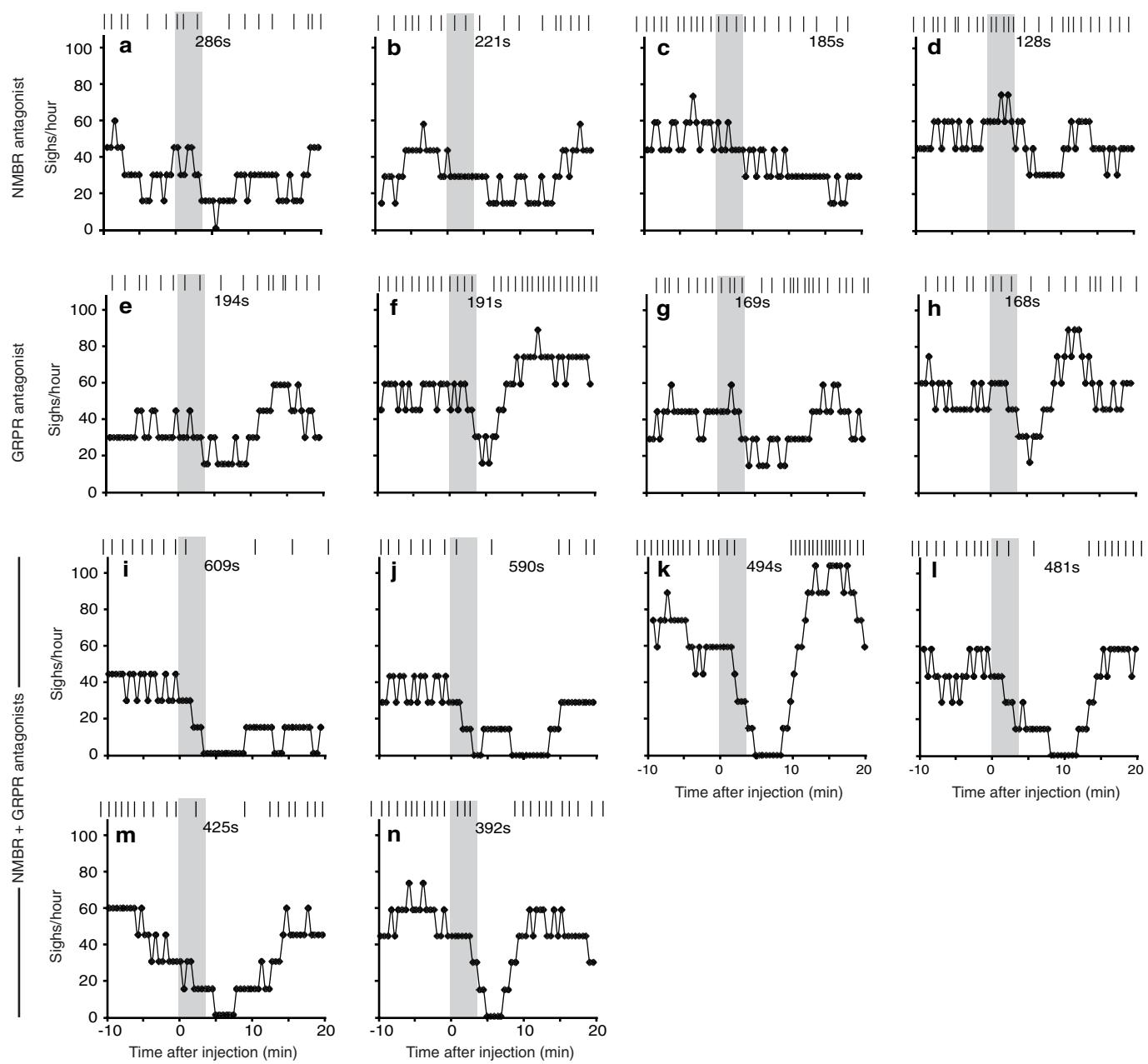
a–e, Raster plot of sighs (upper) and instantaneous sigh rates (lower) before and after NMB injection for the five experiments (a–e) shown in Fig. 2d. **f–j**, Raster plot of sighs (upper) and instantaneous sigh rates (lower) before and after GRP injection for the five experiments (f–j)

shown in Fig. 3c. **k–o**, Raster plot of sighs (upper) and instantaneous sigh rates (lower) before and after NMB/GRP injection for the five experiments (k–o) shown in Fig. 4i. Grey, injection period; arrowhead in raster plots, maximum instantaneous sigh rate; numbers, basal (left) and maximal instantaneous sigh rate (right) and fold induction (in parentheses) after neuropeptide injection.



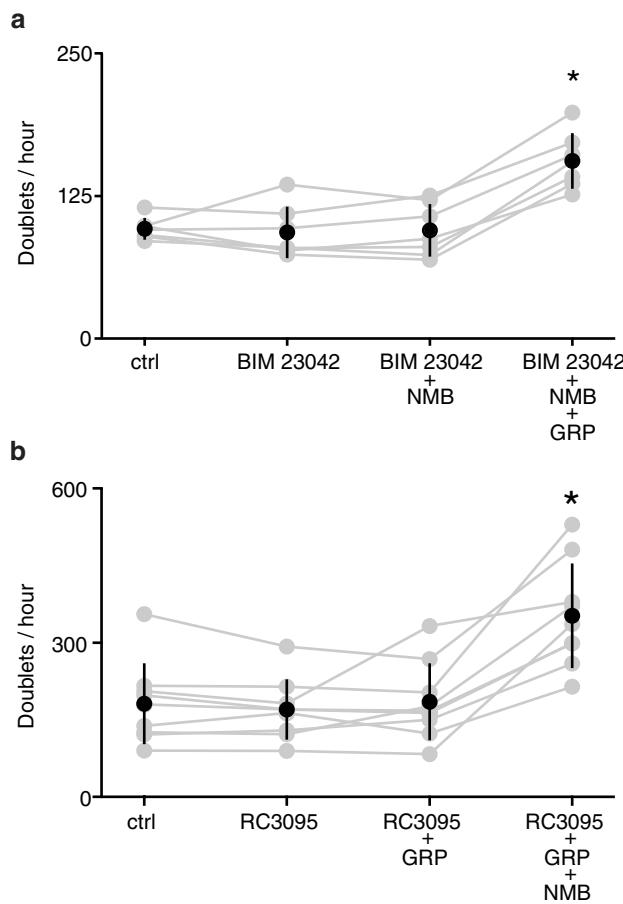
Extended Data Figure 5 | Effect of NMB and GRP on rhythmic activity of preBötC slice. **a**, Neuronal activity trace ($\int XII$, black; $\int preBötC$ population activity, grey) of preBötC slice containing 30 nM NMB, as in Fig. 2e. Note the extreme effect of NMB in which every burst ('breath') in the trace is a doublet ('sigh; asterisk). Bar, 5 s. **b, c**, NMB increases the doublet rate by increasing the fraction of total events that are doublets (**b**) and decreasing the interval following a doublet (**c**). Data are mean \pm s.d., $*P < 0.05$ by paired *t*-test, $n = 7$. **d, e**, GRP also increases the doublet rate by increasing the fraction of total events that are doublets (**d**) and decreasing the interval following a doublet (**e**). Data are mean \pm s.d., $*P < 0.05$ by paired *t*-test, $n = 9$. Note that post-doublet intervals are significantly longer than post-burst intervals under all conditions, consistent with longer post-sigh apneas *in vivo*.

paired *t*-test, $n = 7$. **d, e**, GRP also increases the doublet rate by increasing the fraction of total events that are doublets (**d**) and decreasing the interval following a doublet (**e**). Data are mean \pm s.d., $*P < 0.05$ by paired *t*-test, $n = 9$. Note that post-doublet intervals are significantly longer than post-burst intervals under all conditions, consistent with longer post-sigh apneas *in vivo*.

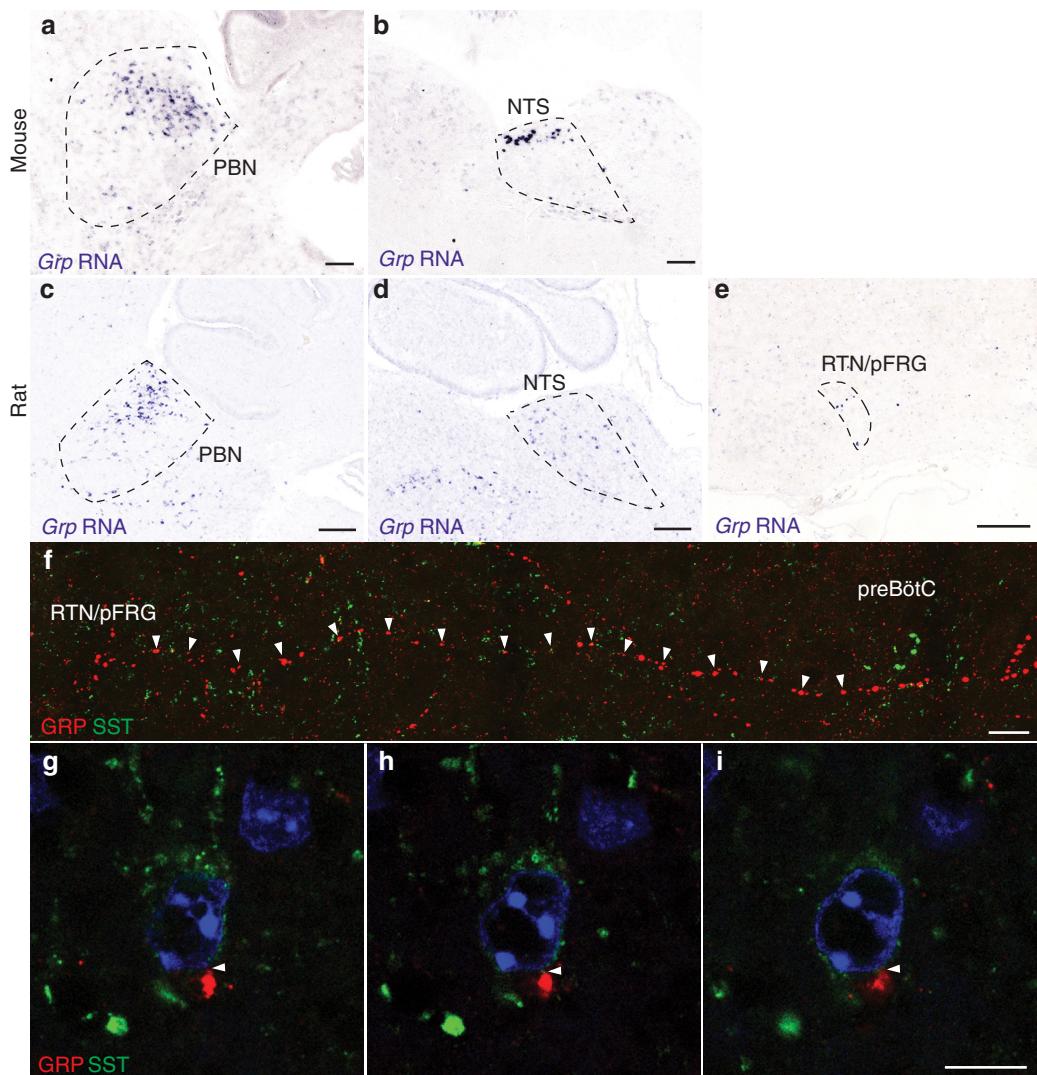


Extended Data Figure 6 | Effects on sighing in individual rats following bilateral injection of BIM23042, RC3095 and BIM23042/RC3095 into preBöTC. **a–d**, Raster plot of sighs (upper) and binned sigh rates (lower; bin size 4 min; slide 30 s) before and after injection of the NMBR antagonist BIM23042 for the four experiments shown in Fig. 2h. **e–h**, Raster plot of sighs (upper) and binned sigh rates (lower; bin size

4 min; slide 30 s) before and after injection of the GRPR antagonist RC3095 for the four experiments shown in Fig. 3f. **i–n**, Raster plot of sighs (upper) and binned sigh rates (lower; bin size 4 min; slide 30 s) before and after BIM23042 and RC3095 injection for the six experiments shown in Fig. 4j. Grey, injection period; numbers, longest inter-sigh intervals (s, seconds) following injection.

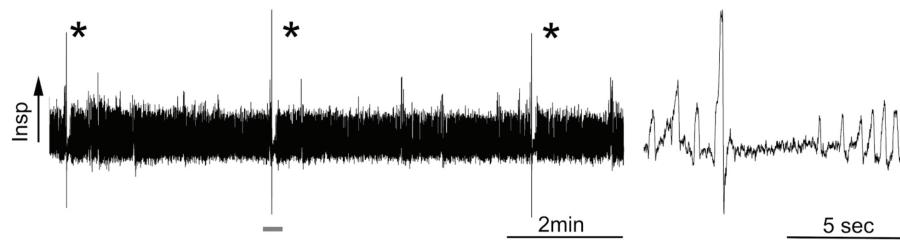


Extended Data Figure 7 | Specificity of antagonists BIM23042 and RC3095 in preBötC slice. **a**, BIM 23042 (100 nM) blocks the effect of NMB (10 nM), but not GRP (3 nM) in preBötC slices. Data are mean \pm s.d., $*P < 0.05$ by paired *t*-test, $n = 7$. **b**, RC3095 (100 nM) shows the opposite specificity, blocking the effect of GRP (3 nM), but not NMB (10 nM). Data are mean \pm s.d., $*P < 0.05$ by paired *t*-test, $n = 9$.

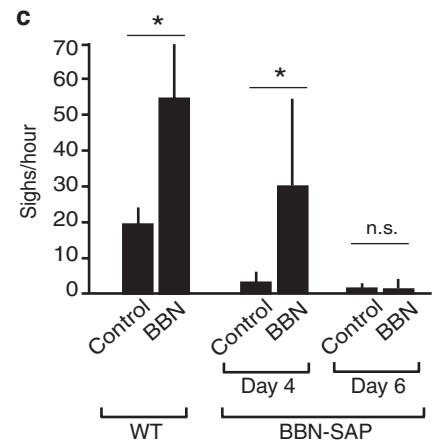


Extended Data Figure 8 | Expression of *Grp* in rodent brain. **a, b,** *In situ* hybridization of mouse brain slices as in Fig. 3a showing expression of *Grp* (purple) in parabrachial nucleus (PBN) (**a**) and nucleus tractus solitarius (NTS) (**b**). Scale bar, 200 μm. **c–e,** *In situ* hybridization of rat brain slices showing expression of *Grp* in PBN (**c**), NTS (**d**), RTN/pFRG (**e**). Scale bar, 200 μm. **f,** Tiled image showing GRP-positive projection (red) from

RTN/pFRG region to preBötC region containing SST-positive neuron (green). Scale bar, 20 μm. **g–i,** Serial confocal optical sections (0.8 μm apart) through mouse preBötC stained for GRP (red) and SST (green) focusing on short segment of GRP-positive projection where a GRP puncta (red) directly abuts (arrowhead) an SST-positive neuron. Scale bar, 10 μm.

a WT**b** BBN-SAP

Extended Data Figure 9 | Effect of bombesin injection on sighing following BBN-SAP-induced ablation of NMBR-expressing and GRPR-expressing preBötC neurons. **a, b**, 10 min plethysmography traces of a control rat (**a**) and a day 5 BBN-SAP injected rat (**b**) during eupneic breathing (left). Indicated parts (10 s) of traces are expanded at right. Note presence of sighs with stereotyped waveform in control rat, and no sighs detectable in BBN-SAP injected rat. **c**, Sigh rate before (control) and after



10 μ g bombesin injection (BBN) into the cisterna magna of rats before BBN-SAP injection (WT) and at day 4 and day 6 after BBN-SAP injection (BBN-SAP) into the preBötC to ablate NMBR-expressing and GRPR-expressing neurons as in Fig. 5a, b. Values shown are mean \pm s.d. (WT, $n = 10$; BBN-SAP, $n = 7$ for day 4 and $n = 5$ for day 6), * $P < 0.05$ by paired t -test; n.s., not significant.

Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 Å resolution

Thi Hoang Duong Nguyen^{1*}, Wojciech P. Galej^{1*}, Xiao-chen Bai¹, Chris Oubridge¹, Andrew J. Newman¹, Sjors H. W. Scheres¹ & Kiyoshi Nagai¹

U4/U6.U5 tri-snRNP represents a substantial part of the spliceosome before activation. A cryo-electron microscopy structure of *Saccharomyces cerevisiae* U4/U6.U5 tri-snRNP at 3.7 Å resolution led to an essentially complete atomic model comprising 30 proteins plus U4/U6 and U5 small nuclear RNAs (snRNAs). The structure reveals striking interweaving interactions of the protein and RNA components, including extended polypeptides penetrating into subunit interfaces. The invariant ACAGAGA sequence of U6 snRNA, which base-pairs with the 5'-splice site during catalytic activation, forms a hairpin stabilized by Dib1 and Prp8 while the adjacent nucleotides interact with the exon binding loop 1 of U5 snRNA. Snu114 harbours GTP, but its putative catalytic histidine is held away from the γ -phosphate by hydrogen bonding to a tyrosine in the amino-terminal domain of Prp8. Mutation of this histidine to alanine has no detectable effect on yeast growth. The structure provides important new insights into the spliceosome activation process leading to the formation of the catalytic centre.

Pre-messenger RNA splicing is catalysed by an intricate molecular machine called the spliceosome and proceeds by a two-step trans-esterification mechanism, analogous to group II intron self-splicing¹. The spliceosome is assembled on pre-mRNA by the ordered addition of small nuclear ribonucleoprotein particles (snRNPs) and numerous proteins, including the nineteen complex (NTC) and the nineteen related (NTR) complex²⁻⁴. Initially U1 and U2 snRNPs recognize the pre-mRNA 5'-splice site (5'SS) and branch point (BP), respectively. Recruitment of U4/U6.U5 tri-snRNP produces the fully assembled but catalytically inactive complex B¹. U1 snRNP is displaced from the 5'SS by Prp28 (ref. 5), the 5'SS pairs with the ACAGAGA sequence in U6 snRNA, and Brr2 helicase unwinds the extensively base-paired U4/U6 snRNAs to release U4 snRNA with its associated proteins^{6,7}. This allows U6 snRNA to base-pair with U2 snRNA generating the group II intron-like catalytic RNA core⁸⁻¹¹. The 2'OH group of the BP adenosine attacks the 5'SS, producing exon1 and lariat intron-exon2 intermediates, and after further remodelling to complex C*, U5 snRNA loop 1 aligns exons 1 and 2 for the second trans-esterification^{12,13}. The spliced mRNA product is released and the residual intron lariat spliceosome (ILS) is disassembled, recycling the snRNPs for subsequent rounds of splicing¹.

Electron microscopic studies of spliceosomes at different stages of the splicing cycle revealed low-resolution pictures of these complexes¹⁴. Taking advantage of the recent revolution in cryo-electron microscopy (cryo-EM) single particle analysis¹⁵, we reported the organization of the proteins, and U5 snRNA and U4/U6 snRNAs in *S. cerevisiae* U4/U6.U5 tri-snRNP based on a cryo-EM map at 5.9 Å (ref. 16). In *Schizosaccharomyces pombe* cell extract, ILS complexes containing U2, U5 and U6 snRNAs accumulate through inefficient disassembly^{17,18}. The structure of this endogenous U2.U6.U5 spliceosomal complex was determined by single particle cryo-EM at 3.6 Å resolution^{19,20}. This was an important breakthrough that revealed the overall architecture of a spliceosomal complex with the striking structures of NTC and NTR^{19,20}. The absence of spliced mRNA and step 2 factors⁴ from this complex¹⁹ confirms that it is the post-splicing ILS¹⁸. The structure also revealed the important features of the well-established group II

intron-like catalytic RNA core⁸⁻¹¹ remaining after spliced mRNA is released²⁰.

Here we present an essentially complete atomic model of *S. cerevisiae* U4/U6.U5 tri-snRNP^{21,22} based on a cryo-EM density map at 3.7 Å overall resolution, revealing the architectural and mechanistic principles of spliceosome activation.

Overall structure

We collected a new data set on a Titan Krios microscope using the Gatan K2 Summit direct electron detector (Methods). The overall resolution of the tri-snRNP map was improved from 5.9 Å to 3.7 Å (Extended Data Fig. 1). Using a modified masked refinement with signal subtraction²³, we obtained more homogeneous 3.6, 3.7 and 4.2 Å reconstructions for the body, foot and head domains, respectively, and improved the resolution of the arm domain from 10 Å to 6–7.5 Å (Extended Data Fig. 2). The new maps enabled us to build a near-complete atomic model of the yeast tri-snRNP containing 30 proteins, U4/U6 and U5 snRNAs (Fig. 1 and Supplementary Information), revealing an amazing web of interactions between components of the complex (Extended Data Fig. 3).

Prp8

A complete atomic model of Prp8 is now built, except for the unstructured N terminus and inter-domain linkers. The α -helix (α RT1) at the N terminus of the reverse transcriptase (RT) domain in the crystal structure²⁴ extends further and forms a helix bundle (HB) with three additional long helices appended to the RT domain (Fig. 2a, b). Residues 108–733 form a predominantly α -helical N-terminal domain. Stems I and II of U5 snRNA are coaxially stacked¹⁶ and an extra variable stem protrudes from the three-way junction (Extended Data Fig. 4). A long, slightly bent C-terminal α -helix (residues 703–735) of the N-terminal domain fits into the minor groove of the co-axially stacked stems I and II, which is tightly harnessed in the major groove by a polypeptide loop (residues 535–543) protruding from the N-terminal domain (Fig. 2c). The conserved loop 1 of U5 snRNA, which aligns the exons during the second trans-esterification reaction^{12,13}, points

¹MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, UK.

*These authors contributed equally to this work.

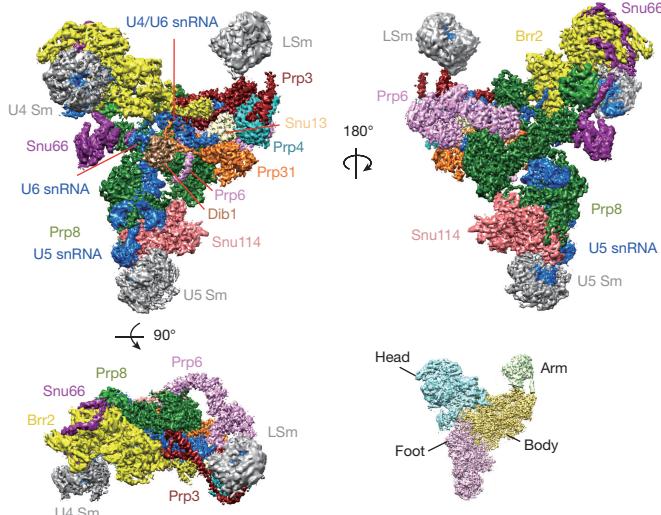


Figure 1 | Three orthogonal views of a near-complete atomic model of the *Saccharomyces cerevisiae* U4/U6.U5 tri-snRNP. Inset shows four sub-domains.

towards the most positively charged and conserved surface of Prp8 in the thumb/linker domain, part of the active site cavity²⁴. The BP+2 nucleotide cross-links in active spliceosomes between Prp8 residues 1585–1598, on the cavity surface (C. M. Norman and A.J.N., unpublished observations). This region is disordered in the Prp8–Aar2 complex²⁴, whereas in U4/U6.U5 tri-snRNP it forms a helix-turn-helix (the α -finger) and contacts U54–U55 of U4 snRNA near the three-way junction (Fig. 2b).

The 5'-stem-loop of U6 snRNA interacts with the N-terminal domain of Prp8 and the adjacent single-stranded region pairs with the exon binding U5 snRNA loop 1 (Fig. 2d). The small highly conserved protein Dib1 (ref. 25) binds to the helix bundle and α -finger of Prp8, and a long polypeptide of Prp31. U6 snRNA forms a short stem-loop, involving part of the ACAGAGA sequence, which is sandwiched between Dib1 and the Prp8 large domain (residues 1648–1653) (Fig. 2d, e; Extended Data Fig. 4a).

Snu114

We built a near complete atomic model of Snu114 comprising five domains (D1–D5) similar to EF-G/EF-2 (refs 26, 27). The relative arrangement of D1–D3 closely resembles that of EF-G/EF-2, whereas D4 and D5 pack more compactly (Fig. 3a). The guanine nucleotide density is consistent with GTP bound via canonical interactions with the surrounding residues (Fig. 3b; Extended Data Figs 3a and 5a–e). In most GTPases the glutamine residue in the switch 2 loop places a water molecule at the γ -phosphate of GTP and hydrolyses the phosphate ester²⁸. As in EF-Tu, EF-G and their eukaryotic counterparts, the catalytic glutamine residue is replaced by histidine in Snu114 (ref. 26) (Extended Data Fig. 5e). In U4/U6.U5 tri-snRNP, His218 is hydrogen-bonded to Tyr403 of Prp8, preventing the His218 side chain from rotating towards the γ -phosphate of GTP and hence keeping the GTPase inactive (Fig. 3c). In EF-G and EF-Tu, GTP is hydrolysed when this histidine is repositioned by a hydrogen-bond with a phosphate in the sarcin-ricin loop of the ribosome^{29,30} (Fig. 3d). The extensive interactions between Snu114 and the N-terminal domain of Prp8 are conserved between U4/U6.U5 tri-snRNP and the *S. pombe* ILS¹⁹ (Extended Data Fig. 5f). The hydrogen bond between His218 of Snu114 and Tyr403 of Prp8 is maintained by the equivalent residues in ILS¹⁹ (Extended Data Fig. 5d). The GTP binding site of Snu114 is at the interface with the N-terminal domain of Prp8, leaving insufficient room for U5 snRNA or any proteins to access the GTPase active site and act like the sarcin-ricin loop^{29,30} or as a GTPase activating protein (GAP)²⁸. Since the structure suggests no obvious mechanism for Snu114 GTPase

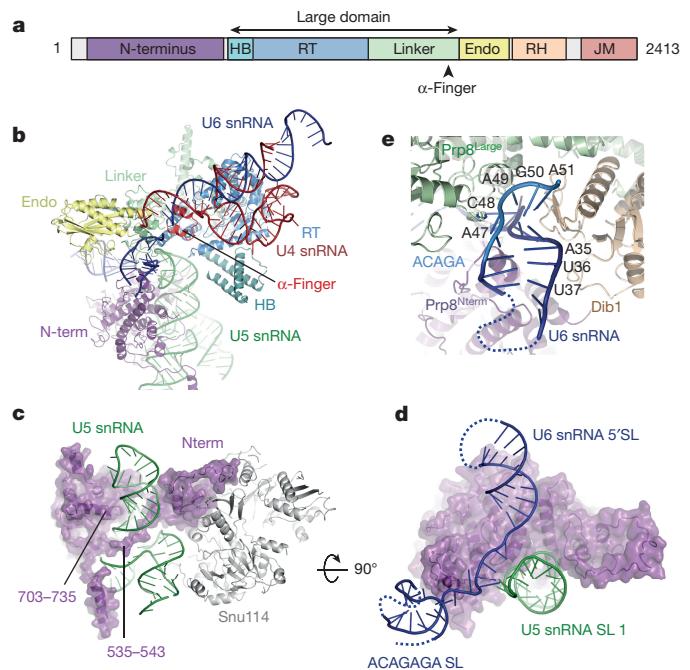


Figure 2 | Prp8 and U4/U6 and U5 snRNAs. a, Domain structure of Prp8²⁴. HB, helix bundle; RT, reverse transcriptase-like; Endo, endonuclease-like; RH, RNase H-like; JM, Jab1/MPN. b, Prp8 makes extensive interactions with U4/U6 and U5 snRNAs. c, The α -helix (residues 703–735) of the N-terminal domain fits into the minor groove of U5 snRNA and an extended polypeptide (residues 535–543) fits into the major groove on the opposite face, harnessing the RNA helix firmly in place. d, Orthogonal view. U5 snRNA loop 1 interacts with the single-stranded region of U6 snRNA adjacent to its 5'-stem-loop. e, The region around the ACAGAGA sequence forms a hairpin and is sandwiched between the large and N-terminal domains of Prp8 and Dib1.

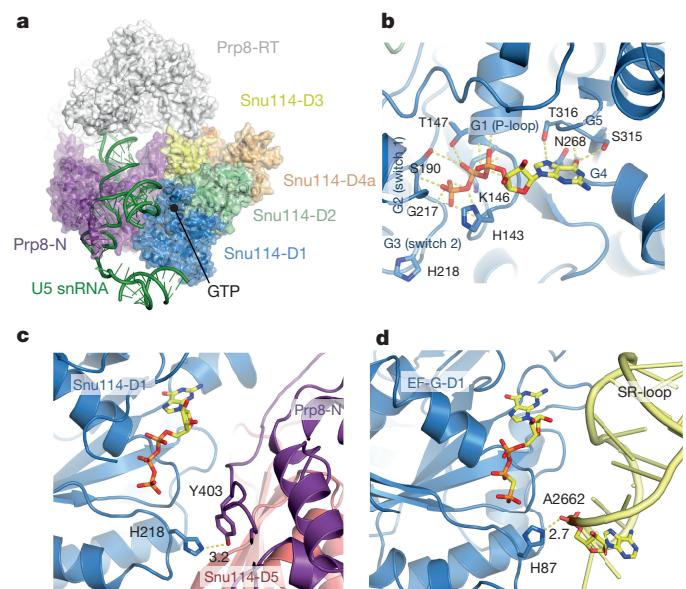


Figure 3 | Snu114 and its interaction with guanine nucleotide, Prp8 and U5 snRNA. a, Snu114, the N-terminal domain of Prp8 and U5 snRNA stems I and II form a stable domain in the foot domain in U4/U6.U5 tri-snRNP. GTP is bound in the GTPase active site at the interface with the Prp8 N-terminal domain. b, Canonical interactions of GTP with surrounding residues in Snu114. c, The catalytic His218, hydrogen bonded to Tyr403 in Prp8, points away from the GTP γ -phosphate. d, Activation of EF-G GTPase upon binding to the sarcin-ricin (SR) loop in the ribosome. His87 moves closer to the γ -phosphate and places a water molecule²⁹.

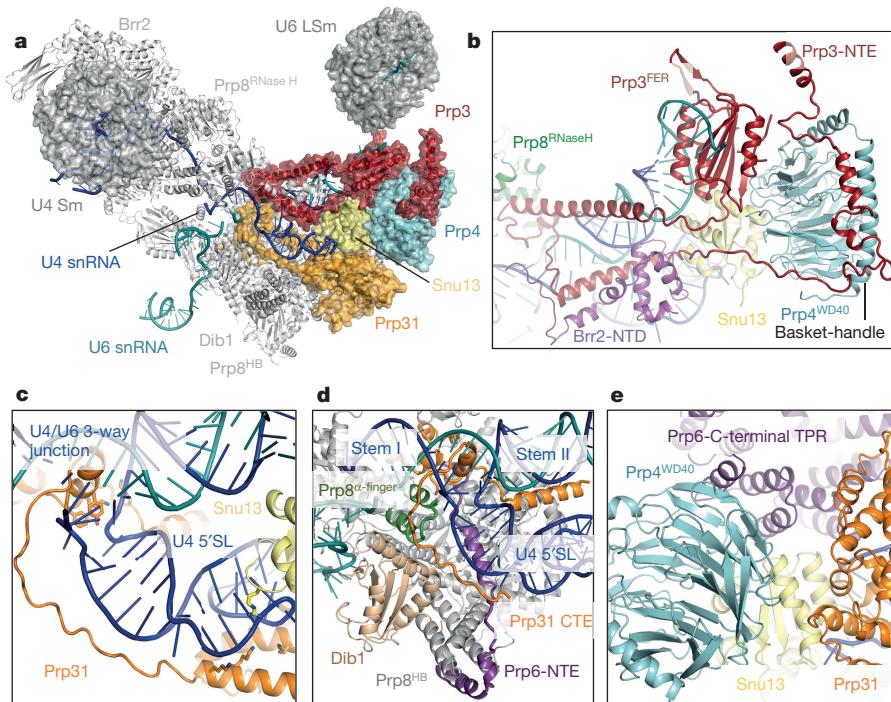


Figure 4 | Interactions of U4/U6 snRNAs with proteins. **a**, Overview of U4/U6 di-snRNP. **b**, The extraordinary structure of Prp3 and its multiple interactions with U4/U6 snRNA, Prp4, Snu13, the RNase H-like domain of Prp8, Brr2 N-terminal domain and the LSm core domain. **c**, The C-terminal region of Prp31 extends along U4 snRNA 5'-stem towards the

activation we investigated the function of Snu114 by mutagenesis. With the His218Arg mutation, yeast shows only a mild temperature-sensitive phenotype, confirming earlier results³¹ (Extended Data Fig. 5g), whereas the equivalent mutation in EF-Tu reduces cognate tRNA-induced GTPase activity 10⁵-fold³². Surprisingly, yeast containing the His218Ala mutant of Snu114 shows no apparent phenotype (Extended Data Fig. 5g), while the equivalent mutation in EF-Tu reduces the rate of GTP hydrolysis more than 10⁶-fold³². Furthermore, mutations of Tyr403 (Tyr403Phe and Tyr403Ala) in Prp8, which hydrogen-bonds with His218 in Snu114, have no apparent phenotype (Extended Data Fig. 5h). These results raise the possibility that Snu114-bound GTP may not be hydrolysed during splicing.

The guanine nucleotide in the post-splicing ILS is interpreted as GDP¹⁹, but its conformation is distinct from that of GDP in other GTPases (Extended Data Fig. 5a). In contrast, the conformation of the Snu114-bound GTP in tri-snRNP superimposes well with GTP or non-hydrolysable GTP analogues in other GTPases (Extended Data Fig. 5c). When we refined our structure of Snu114 with GDP, the resulting GDP conformation is similar to that observed in ILS (Extended Data Fig. 5b). The proposed guanine nucleotide-dependent regulatory role of Snu114 is based on the effects of XDP, XTP and a non-hydrolysable XTP analogue on the XTP binding mutant (D271N) of Snu114 (refs 33, 34). Mutations in the GTPase domain prevent the interaction of Snu114 with Prp8, blocking U5 snRNP assembly³⁵. The observed effect of XDP, XTP and non-hydrolysable XTP may be due to XTP-induced stabilization and association of Snu114 mutants with Prp8.

The U4/U6 di-snRNP

The extensively base-paired U4/U6 snRNAs form a three-way helix junction (Extended Data Figs 4a, b). Snu13, bound to the k-turn motif³⁶, is wedged between the U4 5'-stem-loop and the U4/U6 snRNA helix II and packed against the Prp4 WD40 domain³⁷ (Fig. 4a). Prp3 makes extensive interactions with the Prp4 WD40 domain, the basket handle-like structure and Snu13, and forms a long α -helix sitting in the minor groove of U4/U6 helix II. After forming a short α -helix,

Prp3 folds back to form a long α -helix binding across the major groove of U4/U6 helix II (Fig. 4b; Extended Data Fig. 3d). These latter two Prp3 helices and the connecting loop interact extensively with the RNase H-like domain of Prp8 and Brr2 N-terminal domain. Prp3 further extends to form a ferredoxin-like domain³⁸, which packs against the Prp4 WD40 domain³⁷. Masked classification of the arm domain reveals extra density for the N terminus of Prp3 extending towards the LSm protein ring (Extended Data Fig. 6a, b). The 3'-end of U6 snRNA binds to the central hole of the LSm protein ring while the preceding single-stranded region binds to the ferredoxin-like domain of Prp3 (ref. 38). The Nop and coiled-coil domains of Prp31 interact with Snu13, whereas the k-turn motif of U4 5'-stem-loop is sandwiched between Snu13 and Prp31 (refs 36, 39) (Fig. 4c). The extended polypeptide chain of Prp31 runs between the phosphate backbone of U4 5'-stem and Dib1, and forms a small domain together with Prp6 which is surrounded by the three-way RNA helix junction and the α -finger and helix bundle of Prp8 (Fig. 4d).

The C terminus of the Prp6 TPR repeats⁴⁰ interacts with the Prp4 WD40 domain, Snu13, Prp31 and the tip of U4 5'-stem-loop (Fig. 4e), while an extended N-terminal polypeptide of Prp6 packs against the RNase H-like domain of Prp8 and interacts with the small carboxy-terminal domain of Prp31, the Prp8 α -finger, and U4/U6 snRNA three-way junction and then wraps around the Prp8 helix bundle (Fig. 4d; Extended Data Fig. 3d-f). The numerous interactions that Prp6 makes with U4/U6 snRNP components and Prp8 reflect its importance for tri-snRNP assembly⁴¹.

Brr2

The single-stranded region of U4 snRNA (Extended Data Figs 3b and 4a), extending from stem I, enters the active site of Brr2 N-terminal helicase cassette near the strand-separating β -hairpin and passes through the channel between the RecA1, RecA2, Ratchet and WH domains⁴² (Extended Data Fig. 7a-c). The N-terminal domain (NTD) of Brr2 extends towards U4/U6 stem II and contacts the long helix of Prp3 running along the phosphate backbone of U4 snRNA. Brr2 inserts a loop

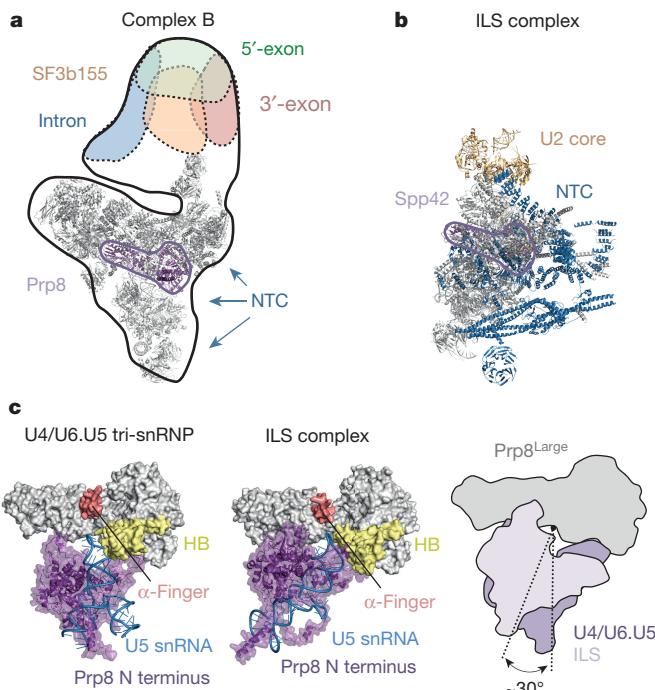


Figure 5 | B complex formation and activation mechanism. **a**, U4/U6.U5 tri-snRNP fits into the EM envelope of human complex B⁴⁵ (reproduced from ref. 45 with permission), showing that U2 snRNP binds near the LSm core domain, Prp6 and Prp3. **b**, Overlay of the Prp8 large domain between tri-snRNP and the ILS¹⁹ shows how NTC/NTR might bind to complex B and interact with U2 snRNP so that U2 snRNP can be passed to the NTC/NTR complex. **c**, A comparison of the tri-snRNP and the ILS¹⁹ structures shows rotation of the foot domain with respect to the Prp8 large domain. Upon rotation, Prp8 residues 602–614 will clash with Dib1 and ACAGAGA helix, causing them to dissociate thus liberating the ACAGAGA sequence to bind the 5'-splice site.

of the NTD into the minor groove of U4/U6 stem II (Extended Data Fig. 7b, d). These interactions may guide U4/U6 stem II during unwinding. Snu13, Prp4 WD40, Prp3 ferredoxin, and Prp31 Nop and coiled-coil domains assemble together while the long α -helices and stretched polypeptide chains of Prp3 and Prp31 extend from these domains and interact with U4/U6 stem II and U4 5'-stem-loop, respectively³⁹. These long α -helices and extended polypeptides may function like elastic bands to accommodate conformational changes and partial strand separation of the U4/U6 duplex as Brr2 translocates along U4 snRNA and unwinds U4/U6 stem I (refs 16, 43). Brr2 forms a stable complex with the Jab1/MPN domain of Prp8 (ref. 42), which is attached to the RNase H-like domain of Prp8 via a long flexible linker, enabling both Brr2 and U4/U6 di-snRNP to detach from the main body of Prp8 during unwinding.

The improved map of the head domain at 4.5–5 Å resolution, obtained by masked refinement, enabled us to build most of the Snu66 structure as poly-Ala chains. Its N-terminal region forms a globular domain that interacts with Prp8 endonuclease-like and Brr2 N-terminal ratchet domains. This is followed by a long helix wedged between Prp8 Jab1/MPN and Brr2 N-terminal HLH domains while its C terminus wraps around Brr2, forming extensive interactions with the Brr2 C-terminal cassette (Extended Data Fig. 7e), fully consistent with yeast two-hybrid and co-immunoprecipitation assays⁴⁴. Interestingly, our global classification approach showed ‘open’ and ‘closed’ conformations of the head and foot domains (Extended Data Fig. 6c–e). In the ‘closed’ conformation, the globular domain of Snu66 contacts the N-terminal domain of Prp8, which in turn interacts with Snu114.

Insight into spliceosome activation

A comparison of U4/U6.U5 tri-snRNP with B⁴⁵ and B Δ U1 complexes⁴⁶ shows that U2 snRNP docks with tri-snRNP where the LSm

complex, Prp3 and Prp6 are located, while U1 snRNP sits on top of U2 snRNP (Fig. 5a). The components of NTC/NTR are also detected by mass spectrometry in complex B³. We compared the structures of our tri-snRNP and the post-splicing ILS¹⁹ by overlaying the large domain of Prp8 together with Snu114 and the U5 core domain. This shows that NTC and NTR can associate with tri-snRNP without clashing and contact U2 snRNP (Fig. 5a, b). In complex B, U2 snRNP interacts with U4/U6.U5 tri-snRNP⁴⁶, but when NTC and NTR dock with tri-snRNP, U2 snRNP is passed to NTC and NTR, and U2 Sm domain and U2B''/U2A' complex associate with Aquarius(Cwf11), Syf1(Cwf3) and Isy1(Cwf12)⁴⁷ as revealed in the *S. pombe* ILS¹⁹ (*S. pombe* protein names are shown in parentheses).

The *S. cerevisiae* U4/U6.U5 tri-snRNP and *S. pombe* ILS structures reveal that the foot domains of the two structures, containing the Prp8 N-terminal domain, U5 snRNA stem-loop 1 and Snu114, superpose very well showing that they form a stable structural unit (Extended Data Fig. 5f). Overlay of their Prp8 large domains shows that the foot domain rotates as a rigid body by 30° between the two structures, causing U5 loop 1 to move closer towards the Prp8 α -finger in the post-splicing ILS¹⁹ (Fig. 5c). NTC forms extensive interfaces with both the N-terminal and large domains of Prp8, hence the rotation of the foot domain may be caused by NTC. When the foot domain of the U4/U6.U5 tri-snRNP structure rotates by 30° (as in the post-splicing ILS) Prp8 residues 602–614 clash with Dib1 and the ACAGAGA helix, forcing Dib1 to dissociate from the large domain of Prp8 and liberating the ACAGAGA sequence to bind the 5'SS. It is known from the U4 snRNA cs1 mutation and its suppressor in U6 snRNA (U6-Dup)⁴⁸ that the pairing between 5'SS and the ACAGAGA sequence is a checkpoint for the unwinding of the U4/U6 snRNA duplex by Brr2. Thus, conformational toggling of the Prp8 N-terminal domain could couple 5'SS recognition to U4/U6 unwinding. Suppressors of the U4-cs1 mutation⁴⁹ suggest the allosteric changes required for Brr2 activation. Interestingly, these suppressors form four clusters in Prp8: one at the interface between the RT domain and D3 of Snu114, one at the interface between the helix bundle and Prp31, one on the surface of the endonuclease-like domain near the ACAGAGA hairpin, and one on the surface of the N-terminal domain where the 5'-stem-loop of U6 snRNA binds, near the interface with Snu66 that undergoes a transition between the open and closed forms (Extended Data Fig. 6c–e). It is possible that NTC/NTR play important roles in inducing allosteric changes that trigger the unwinding of U4/U6 snRNA by Brr2 (ref. 50).

The cryo-EM structures of *S. cerevisiae* U4/U6.U5 tri-snRNP and *S. pombe* ILS¹⁹ have provided a wealth of new information about the architecture and conformational changes of these spliceosomal assemblies. Functional studies based on these new structural insights should greatly enhance our understanding of spliceosome activation and catalysis.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 3 November; accepted 18 December 2015.

Published online 1 February; corrected online 17 February 2016
(see full-text HTML version for details).

1. Will, C. L. & Lührmann, R. Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* **3**, a003707 (2011).
2. Chen, C. H. *et al.* Functional and physical interactions between components of the Prp19-associated complex. *Nucleic Acids Res.* **30**, 1029–1037 (2002).
3. Fabrizio, P. *et al.* The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome. *Mol. Cell* **36**, 593–608 (2009).
4. Ohr, T. *et al.* Molecular dissection of step 2 catalysis of yeast pre-mRNA splicing investigated in a purified system. *RNA* **19**, 902–915 (2013).
5. Staley, J. P. & Guthrie, C. An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. *Mol. Cell* **3**, 55–64 (1999).
6. Raghunathan, P. L. & Guthrie, C. RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. *Curr. Biol.* **8**, 847–855 (1998).
7. Laggerbauer, B., Achsel, T. & Lührmann, R. The human U5-200KD DEIH-box protein unwinds U4/U6 RNA duplexes *in vitro*. *Proc. Natl. Acad. Sci. USA* **95**, 4188–4192 (1998).

8. Fica, S. M. *et al.* RNA catalyses nuclear pre-mRNA splicing. *Nature* **503**, 229–234 (2013).
9. Madhani, H. D. & Guthrie, C. Dynamic RNA–RNA interactions in the spliceosome. *Annu. Rev. Genet.* **28**, 1–26 (1994).
10. Toor, N., Keating, K. S., Taylor, S. D. & Pyle, A. M. Crystal structure of a self-spliced group II intron. *Science* **320**, 77–82 (2008).
11. Fica, S. M., Mefford, M. A., Piccirilli, J. A. & Staley, J. P. Evidence for a group II intron-like catalytic triplex in the spliceosome. *Nature Struct. Mol. Biol.* **21**, 464–471 (2014).
12. Newman, A. J. & Norman, C. U5 snRNA interacts with exon sequences at 5' and 3' splice sites. *Cell* **68**, 743–754 (1992).
13. Sontheimer, E. J. & Steitz, J. A. The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science* **262**, 1989–1996 (1993).
14. Stark, H. & Lührmann, R. Cryo-electron microscopy of spliceosomal components. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 435–457 (2006).
15. Bai, X. C., McMullan, G. & Scheres, S. H. How cryo-EM is revolutionizing structural biology. *Trends Biochem. Sci.* **40**, 49–57 (2015).
16. Nguyen, T. H. *et al.* The architecture of the spliceosomal U4/U6.U5 tri-snRNP. *Nature* **523**, 47–52 (2015).
17. Ohi, M. D., Ren, L., Wall, J. S., Gould, K. L. & Walz, T. Structural characterization of the fission yeast U5.U2/U6 spliceosome complex. *Proc. Natl. Acad. Sci. USA* **104**, 3195–3200 (2007).
18. Chen, W. *et al.* Endogenous U2–U5–U6 snRNA complexes in *S. pombe* are intron lariat spliceosomes. *RNA* **20**, 308–320 (2014).
19. Yan, C. *et al.* Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science* **349**, 1182–1191 (2015).
20. Hang, J., Wan, R., Yan, C. & Shi, Y. Structural basis of pre-mRNA splicing. *Science* **349**, 1191–1198 (2015).
21. Stevens, S. W. *et al.* Biochemical and genetic analyses of the U5, U6, and U4/U6 x U5 small nuclear ribonucleoproteins from *Saccharomyces cerevisiae*. *RNA* **7**, 1543–1553 (2001).
22. Gottschalk, A. *et al.* Identification by mass spectrometry and functional analysis of novel proteins of the yeast [U4/U6.U5] tri-snRNP. *EMBO J.* **18**, 4535–4548 (1999).
23. Bai, X.-C., Rajendra, E., Yang, G., Shi, Y. & Scheres, S. H. W. Sampling the conformational space of the catalytic subunit of human γ -secretase. *eLife* **4**, e11182 (2015).
24. Galej, W. P., Oubridge, C., Newman, A. J. & Nagai, K. Crystal structure of Prp8 reveals active site cavity of the spliceosome. *Nature* **493**, 638–643 (2013).
25. Reuter, K., Nottrott, S., Fabrizio, P., Lührmann, R. & Ficner, R. Identification, characterization and crystal structure analysis of the human spliceosomal U5 snRNP-specific 15 kD protein. *J. Mol. Biol.* **294**, 515–525 (1999).
26. Fabrizio, P., Laggerbauer, B., Lauber, J., Lane, W. S. & Lührmann, R. An evolutionarily conserved U5 snRNP-specific protein is a GTP-binding factor closely related to the ribosomal translocase EF-2. *EMBO J.* **16**, 4092–4106 (1997).
27. Jørgensen, R. *et al.* Two crystal structures demonstrate large conformational changes in the eukaryotic ribosomal translocase. *Nature Struct. Biol.* **10**, 379–385 (2003).
28. Wittinghofer, A. & Vetter, I. R. Structure–function relationships of the G domain, a canonical switch motif. *Annu. Rev. Biochem.* **80**, 943–971 (2011).
29. Tourigny, D. S., Fernández, I. S., Kelley, A. C. & Ramakrishnan, V. Elongation factor G bound to the ribosome in an intermediate state of translocation. *Science* **340**, 1235490 (2013).
30. Lin, J., Gagnon, M. G., Bulkley, D. & Steitz, T. A. Conformational changes of elongation factor G on the ribosome during tRNA translocation. *Cell* **160**, 219–227 (2015).
31. Brenner, T. J. & Guthrie, C. Genetic analysis reveals a role for the C terminus of the *Saccharomyces cerevisiae* GTPase Snu114 during spliceosome activation. *Genetics* **170**, 1063–1080 (2005).
32. Maracci, C., Peske, F., Dannies, E., Pohl, C. & Rodnina, M. V. Ribosome-induced tuning of GTP hydrolysis by a translational GTPase. *Proc. Natl. Acad. Sci. USA* **111**, 14418–14423 (2014).
33. Small, E. C., Leggett, S. R., Winans, A. A. & Staley, J. P. The EF-G-like GTPase Snu114p regulates spliceosome dynamics mediated by Brr2p, a DExD/H box ATPase. *Mol. Cell* **23**, 389–399 (2006).
34. Bartels, C., Urlaub, H., Lührmann, R. & Fabrizio, P. Mutagenesis suggests several roles of Snu114p in pre-mRNA splicing. *J. Biol. Chem.* **278**, 28324–28334 (2003).
35. Brenner, T. J. & Guthrie, C. Assembly of Snu114 into U5 snRNP requires Prp8 and a functional GTPase domain. *RNA* **12**, 862–871 (2006).
36. Liu, S. *et al.* Binding of the human Prp31 Nop domain to a composite RNA–protein platform in U4 snRNP. *Science* **316**, 115–120 (2007).
37. Ayadi, L. *et al.* Functional and structural characterization of the Prp3 binding domain of the yeast Prp4 splicing factor. *J. Mol. Biol.* **284**, 673–687 (1998).
38. Liu, S. *et al.* A composite double-/single-stranded RNA-binding region in protein Prp3 supports tri-snRNP stability and splicing. *eLife* **4**, e07320 (2015).
39. Nottrott, S., Urlaub, H. & Lührmann, R. Hierarchical, clustered protein interactions with U4/U6 snRNA: a biochemical role for U4/U6 proteins. *EMBO J.* **21**, 5527–5538 (2002).
40. Makarov, E. M., Makarova, O. V., Achsel, T. & Lührmann, R. The human homologue of the yeast splicing factor prp6p contains multiple TPR elements and is stably associated with the U5 snRNP via protein–protein interactions. *J. Mol. Biol.* **298**, 567–575 (2000).
41. Galisson, F. & Legrain, P. The biochemical defects of *prp4-1* and *prp6-1* yeast splicing mutants reveal that the PRP6 protein is required for the accumulation of the [U4/U6.U5] tri-snRNP. *Nucleic Acids Res.* **21**, 1555–1562 (1993).
42. Nguyen, T. H. D. *et al.* Structural basis of Brr2–Prp8 interactions and implications for U5 snRNP biogenesis and the spliceosome active site. *Structure* **21**, 910–919 (2013).
43. Hahn, D., Kudla, G., Tollervey, D. & Beggs, J. D. Brr2p-mediated conformational rearrangements in the spliceosome during activation and substrate repositioning. *Genes Dev.* **26**, 2408–2421 (2012).
44. van Nues, R. W. & Beggs, J. D. Functional contacts with a range of splicing proteins suggest a central role for Brr2p in the dynamic control of the order of events in spliceosomes of *Saccharomyces cerevisiae*. *Genetics* **157**, 1451–1467 (2001).
45. Wolf, E. *et al.* Exon, intron and splice site locations in the spliceosomal B complex. *EMBO J.* **28**, 2283–2292 (2009).
46. Boehringer, D. *et al.* Three-dimensional structure of a pre-catalytic human spliceosomal complex B. *Nature Struct. Mol. Biol.* **11**, 463–468 (2004).
47. De, I. *et al.* The RNA helicase Aquarius exhibits structural adaptations mediating its recruitment to spliceosomes. *Nature Struct. Mol. Biol.* **22**, 138–144 (2015).
48. Li, Z. & Brow, D. A. Spontaneous duplication in U6 spliceosomal RNA uncouples the early and late functions of the ACAGA element *in vivo*. *RNA* **2**, 879–894 (1996).
49. Kuhn, A. N. & Brow, D. A. Suppressors of a cold-sensitive mutation in yeast U4 RNA define five domains in the splicing factor Prp8 that influence spliceosome activation. *Genetics* **155**, 1667–1682 (2000).
50. Chan, S. P. & Cheng, S. C. The Prp19-associated complex is required for specifying interactions of U5 and U6 with pre-mRNA during spliceosome activation. *J. Biol. Chem.* **280**, 31190–31199 (2005).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank C. Savva, S. Chen, G. McMullan, J. Grimmett and T. Darling for running the electron microscopy and computing facilities, A. Brown, P. Emsley, G. Murshudov for advice and help with model building and refinement, R. O'Keefe for the Δ Snu114 yeast strain, the members of the spliceosome group for help and advice throughout the project and R. Leiro for help with data processing. We thank S. Fica for critical reading of the manuscript and J. Löwe, V. Ramakrishnan, and R. Henderson for their continuing support and encouragements. The project was supported by the Medical Research Council (MC_U105184330 to K.N. and MC_UP_A025_1013 to S.H.W.S.).

Author Contributions T.H.D.N. purified yeast tri-snRNP and prepared EM grids, T.H.D.N., W.P.G. and X.-C.B. collected all EM images. T.H.D.N. processed data and calculated the maps with the help of X.-C.B and S.H.W.S.; T.H.D.N., W.P.G. and C.O. built a model into the map and refined the structure. T.H.D.N. performed mutagenesis experiments with the help of A.J.N.; T.H.D.N. and W.P.G. prepared all illustrations. K.N. initiated and orchestrated the project. T.H.D.N., W.P.G. and K.N. analysed the structure and wrote the paper with invaluable contributions from all other authors.

Author Information The cryo-EM maps have been deposited in the Electron Microscopy Data Bank with accession codes EMD-8006, EMD-8007, EMD-8008, EMD-8009, EMD-8010, EMD-8011, EMD-8012, EMD-8013 and EMD-8014. The coordinates of the atomic models have been deposited in the Protein Data Bank under accession codes 5GAN (overall), 5GAP (body domain), 5GAO (head domain) and 5GAM (foot domain). Reprints and permissions information are available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.H.D.N. (knguyen@mrc-lmb.cam.ac.uk), W.P.G (wgalej@mrc-lmb.cam.ac.uk) or K.N. (kn@mrc-lmb.cam.ac.uk).

METHODS

Statistics. No statistical methods were used to predetermine sample size.

Sample preparation. Tri-snRNP sample was prepared as described in our published protocol¹⁶.

Electron microscopy. Aliquots of 3.5 µl of purified yeast tri-snRNP were applied to Quantifoil Cu R1.2/1.3, 400 mesh grids, which were coated with 6 nm-thick homemade carbon film and glow-discharged in *N*-amylamine. The grids were blotted for 2 s at 4 °C, plunged into liquid ethane by an FEI Vitrobot MKIII at 100% humidity and loaded onto a Titan Krios transmission electron microscope operated at 300 kV. Zero-loss-energy images were collected manually on a Gatan K2-Summit detector in super-resolution counting mode at a calibrated magnification of 35,714× (pixel size of 1.43 Å) and a dose rate of ~2.5 electrons per Å² per second (Extended Data Fig. 1a). We used a slit width of 20 eV on a GIF Quantum energy filter. Each image was exposed for a total of 16 s and dose-fractionated into 20 movie frames. A defocus range of 0.5–3.5 µm was used.

Image processing. MOTIONCORR⁵¹ was used for whole-image drift correction of the movie frames of each micrograph, and contrast transfer function (CTF) parameters of the corrected micrographs were estimated using CTFFIND4 (ref. 52). All subsequent processing steps were done using RELION⁵³ unless otherwise stated. A subset of ~5,000 particles was picked manually, extracted using a 380² pixel box and subjected to reference-free 2D classification. Some of the resulting 2D class averages were low-pass filtered to 20 Å and used as references for automatic particle picking of the whole data set of 2,477 micrographs. The automatically picked particles were screened manually to remove false positives, aggregation and ice contamination, resulting in an initial set of 473,827 particles for reference-free 2D classification. We selected 438,602 particles from good 2D classes for the 3D classification (Extended Data Fig. 1b, c), which was run for 25 iterations, using an angular sampling of 7.5°, a regularisation parameter *T* of 4 and a 60 Å low-pass filtered initial model from our previous reconstruction¹⁶. A subset of 140,155 particles was selected for the first 3D auto-refinement. Particle-based beam-induced motion correction and radiation-damage weighing (particle polishing) were performed on these particles⁵⁴. Auto-refinement of the polished particles resulted in a reconstruction at 3.7 Å overall resolution with an estimated angular accuracy of 1.1°.

Local resolution analysis by Resmap⁵⁵ showed a range of resolution from 3.0 Å in the core to 10 Å in the arm domain and part of the head domain, indicating conformational heterogeneity within the complex. As previously observed, the four domains of the structure, particularly the head and arm domains, are flexible in our structure. We employed two classification/refinement approaches: a local approach to improve the local resolution of the domains and a global approach to allow global conformations of the domains relative to one another to be observed (Extended Data Fig. 1c). For the local approach, we used a masked refinement procedure with signal subtraction for each of the head, body and foot domains²³ and a masked classification with signal subtraction followed by a masked refinement for the most flexible arm domain²³. Each of the four domains only makes up a third or less of the total mass of the complex. For each domain, we subtracted projections from the remaining three domains of the reconstruction in the experimental particle images using the relative orientation of each experimental image from the last auto-refinement run of all the polished particles. This resulted in four sets of new experimental particle images that only have signal from the domain of interest. For the body, foot and head domains the subtracted experimental images were used in 3D auto-refinement with a soft mask for that domain, yielding 3.6, 3.7 and 4.2 Å reconstructions for the body, foot and head domains, respectively (Extended Data Figs 1b, 2a–c). The arm domain is too small for masked refinement. Thus we performed 3D classification on the subtracted experimental images with a mask around the arm domain and no alignments. We selected three classes with 23,760, 26,367 and 24,627 particles each with three distinct conformations for the arm domain. Since the arm domain is too small for accurate alignments of the particles, we refined each of these classes together with the body domain using a new set of modified experimental particle images that included both the arm and body domains by the same subtraction method used previously (Extended Data Figs 1c and 6a, b). Class 1 with 23,760 particles yielded a 4.6 Å overall resolution for both body and arm domains and 6.2 Å resolution for the arm domain alone. Class 2 with 26,363 particles yielded a 4.5 Å overall resolution for both body and arm domains and 7.5 Å resolution for the arm domain alone. Class 3 with 24,627 particles yielded a 4.4 Å overall resolution for both body and arm domains and 6.2 Å resolution for the arm domain alone.

For the global classification/refinement approach, we performed 3D classification of the polished particles for the whole complex with a finer angular sampling of 1.8° and local angular search range of 10°. Two of the sub-classes of 48,945 and 36,824 particles had significantly better angular accuracies and gave 4.2 Å and 4.3 Å reconstructions, respectively, after auto-refinement with more homogeneous

conformations of the head and foot domains. We observed distinct “open” and “closed” conformations for the head and foot domains (Extended Data Fig. 6c–e).

All reported resolutions are based on the gold-standard Fourier shell correlation (FSC) = 0.143 criterion⁵⁶. FSC curves were calculated using soft spherical masks and high-resolution noise substitution was used to correct for convolution effects of the masks on the FSC curves⁵⁷. Prior to visualization, all maps were corrected for the modulation transfer function of the detector. Local resolution was estimated using Resmap⁵⁵.

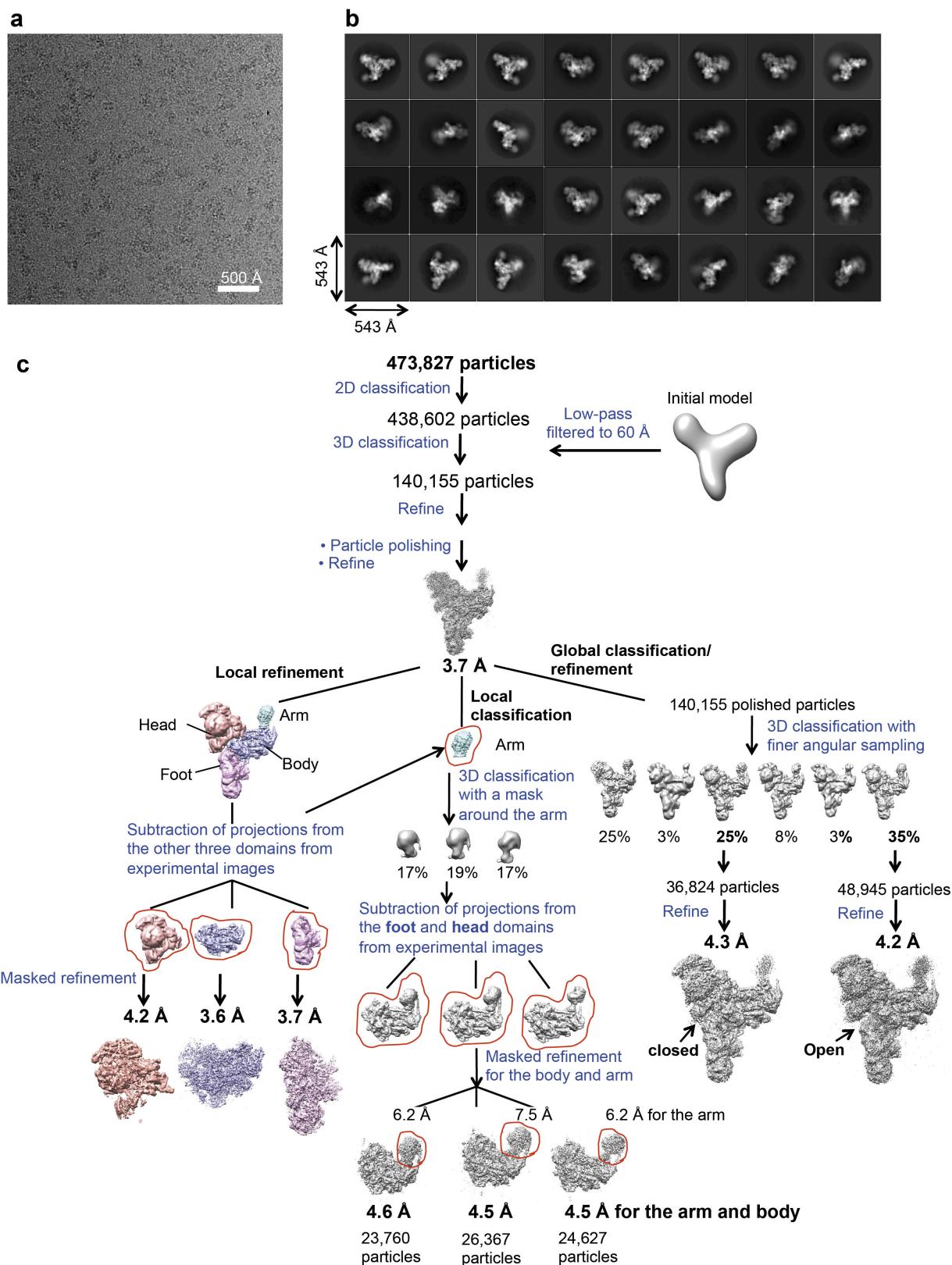
Model building. The maps resulting from local masked refinements were first used for *de novo* model building using our previous protein placements¹⁵ because they have the best local resolution for each of the domains separately. All model building was performed in COOT⁵⁸. In our medium resolution structure, except for Brr2^{442–2163}, Prp8^{885–2413}, Snu13 and LSm proteins whose yeast structures are available, all other proteins are either human, homology models or idealized poly-Ala helices and only double-stranded RNA helices were modelled. Recently the structure of the ferredoxin domain of yeast Prp3 (residues 335–467) became available³⁸, which replaced our homology model of this domain. We rebuilt and extended the yeast Brr2^{442–2163}, Snu13, Prp8^{885–2413} and Prp3^{335–467} and all remaining components were built *de novo* first into the masked refinement maps and rigid-body fitted into the overall 3.7 Å map. We identified a previously unassigned density as that of Snu66 based on previous yeast two-hybrid studies⁴⁴ and its interacting proteins in our structure. The LSm proteins were rigid-body fitted into the overall map and the improved maps of the 3 classes from masked classification and refinement. Extended Data Table 1 summarizes all modelled components of the structure. The model was refined using REFMAC 5.8 (ref. 59) with secondary structure restraints provided by PROSMART⁶⁰ and RNA base-pair and stacking restraints provided by LIBG⁶¹. We first performed model refinement for the Body, Foot and Head domains separately against the corresponding masked refined maps (Extended Data Table 2a). The subunits of these three refined models were rigid-body fitted into the masked overall map. To resolve the possible clashes in the domain interfaces, we refined this overall model against the overall map. Cross-validation of two half maps defined a REFMAC refinement weight of 0.001. The Xmipp package⁶² was used to calculate FSC model versus map. FSC curves of model versus map were calculated for the maps of the body, foot and head domains, which were used for model building and refinement of the structure and also the overall map (Extended Data Fig. 2d–g). Extended Data Table 2 summarizes refinement statistics for the overall structure and the domain structures and the deposited maps and their associated coordinates.

Map visualization. Maps were visualized in Chimera⁶³ and all figures were prepared using either Pymol (<http://www.pymol.org>) or Chimera.

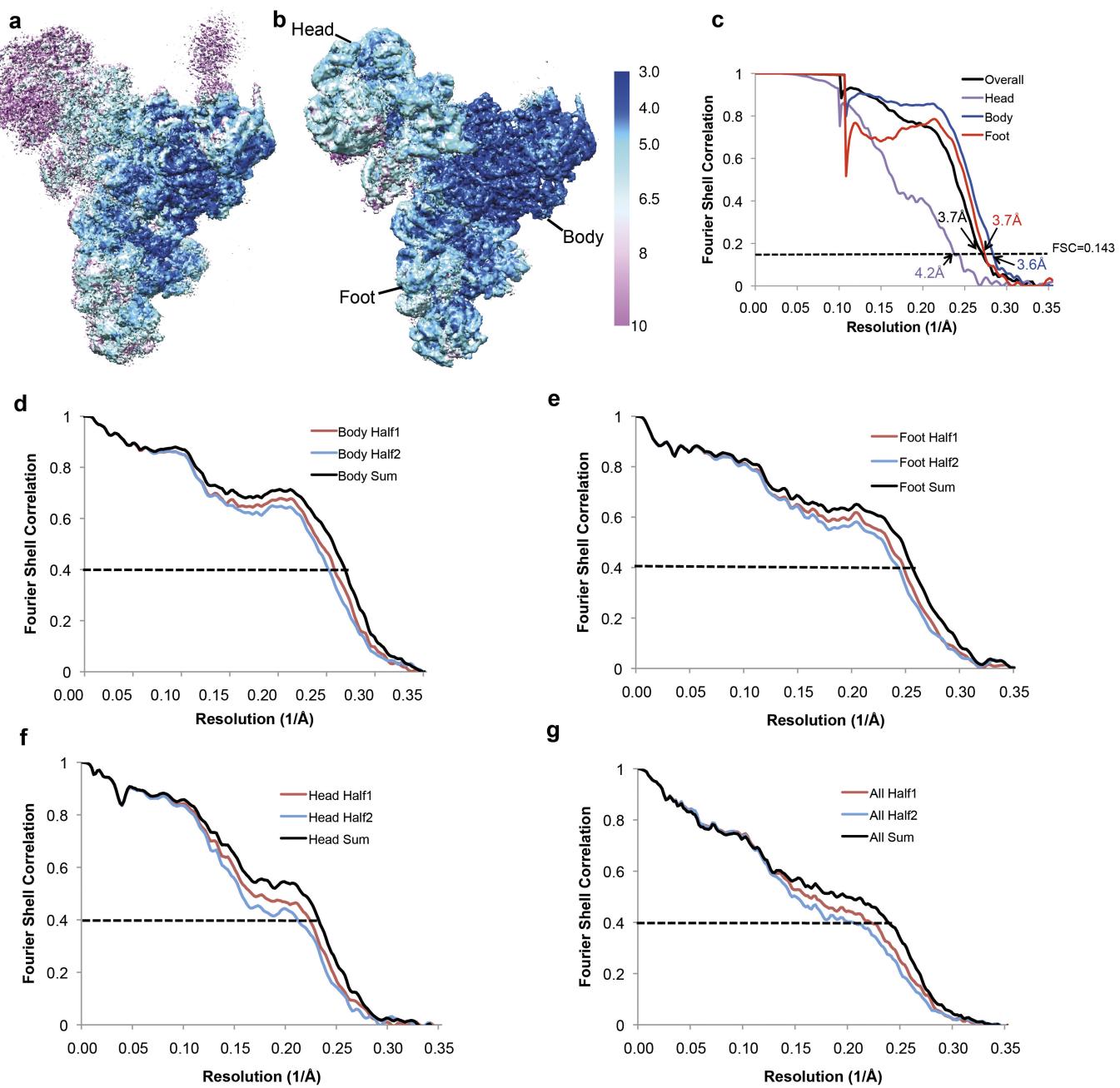
Plasmid Shuffling. Mutations were introduced into *PRP8* and *SNU114* genes by the *dut*[–] *ung*[–] methods⁶⁴. The viability of *Prp8* and *Snu114* mutants was assessed by plasmid shuffling analysis. The *Prp8* deletion strain SC261Δ8B1 (ref. 65) carrying wild type *PRP8* on pRS316 (*URA3*, centromeric replication origin) was transformed with mutant *Prp8* genes on pRS314 (*TRP1*, centromeric replication origin) and transformants were selected on plates lacking tryptophan. The *Snu114* deletion strain YSNU114KO1 (ref. 66) carrying wild type *SNU114* on pRS416 (*URA3*, centromeric replication origin) was transformed with mutant *Snu114* genes on pRS413 (*HIS3*, centromeric replication origin) and transformants were selected on plates lacking histidine. Trp⁺ and His⁺ cells were transferred onto plates containing 5-fluoro-orotic acid (5-FOA), to test cell growth at 30 °C after loss of the *URA3*-marked plasmid. Plasmids were rescued from the 5-FOA-resistant strains and sequenced to confirm the presence of the appropriate mutation, and cell growth was assessed on YEPD plates incubated at various temperatures.

- Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods* **10**, 584–590 (2013).
- Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
- Scheres, S. H. A Bayesian view on cryo-EM structure determination. *J. Mol. Biol.* **415**, 406–418 (2012).
- Scheres, S. H. Beam-induced motion correction for sub-megadalton cryo-EM particles. *eLife* **3**, e03665 (2014).
- Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nature Methods* **11**, 63–65 (2014).
- Scheres, S. H. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nature Methods* **9**, 853–854 (2012).
- Chen, S. *et al.* High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24–35 (2013).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).

59. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355–367 (2011).
60. Nicholls, R. A., Fischer, M., McNicholas, S. & Murshudov, G. N. Conformation-independent structural comparison of macromolecules with ProSMART. *Acta Crystallogr. D* **70**, 2487–2499 (2014).
61. Brown, A. *et al.* Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta Crystallogr. D* **71**, 136–153 (2015).
62. Scheres, S. H. W., Núñez-Ramírez, R., Sorzano, C. O. S., Carazo, J. M. & Marabini, R. Image processing for electron microscopy single-particle analysis using XMIPP. *Nature Protocols* **3**, 977–990 (2008).
63. Goddard, T. D., Huang, C. C. & Ferrin, T. E. Visualizing density maps with UCSF Chimera. *J. Struct. Biol.* **157**, 281–287 (2007).
64. Kunkel, T. A. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl Acad. Sci. USA* **82**, 488–492 (1985).
65. Turner, I. A., Norman, C. M., Churcher, M. J. & Newman, A. J. Dissection of Prp8 protein defines multiple interactions with crucial RNA sequences in the catalytic core of the spliceosome. *RNA* **12**, 375–386 (2006).
66. Frazer, L. N., Lovell, S. C. & O'Keefe, R. T. Analysis of synthetic lethality reveals genetic interactions between the GTPase Snu114p and snRNAs in the catalytic core of the *Saccharomyces cerevisiae* spliceosome. *Genetics* **183**, 497–515 (2009).

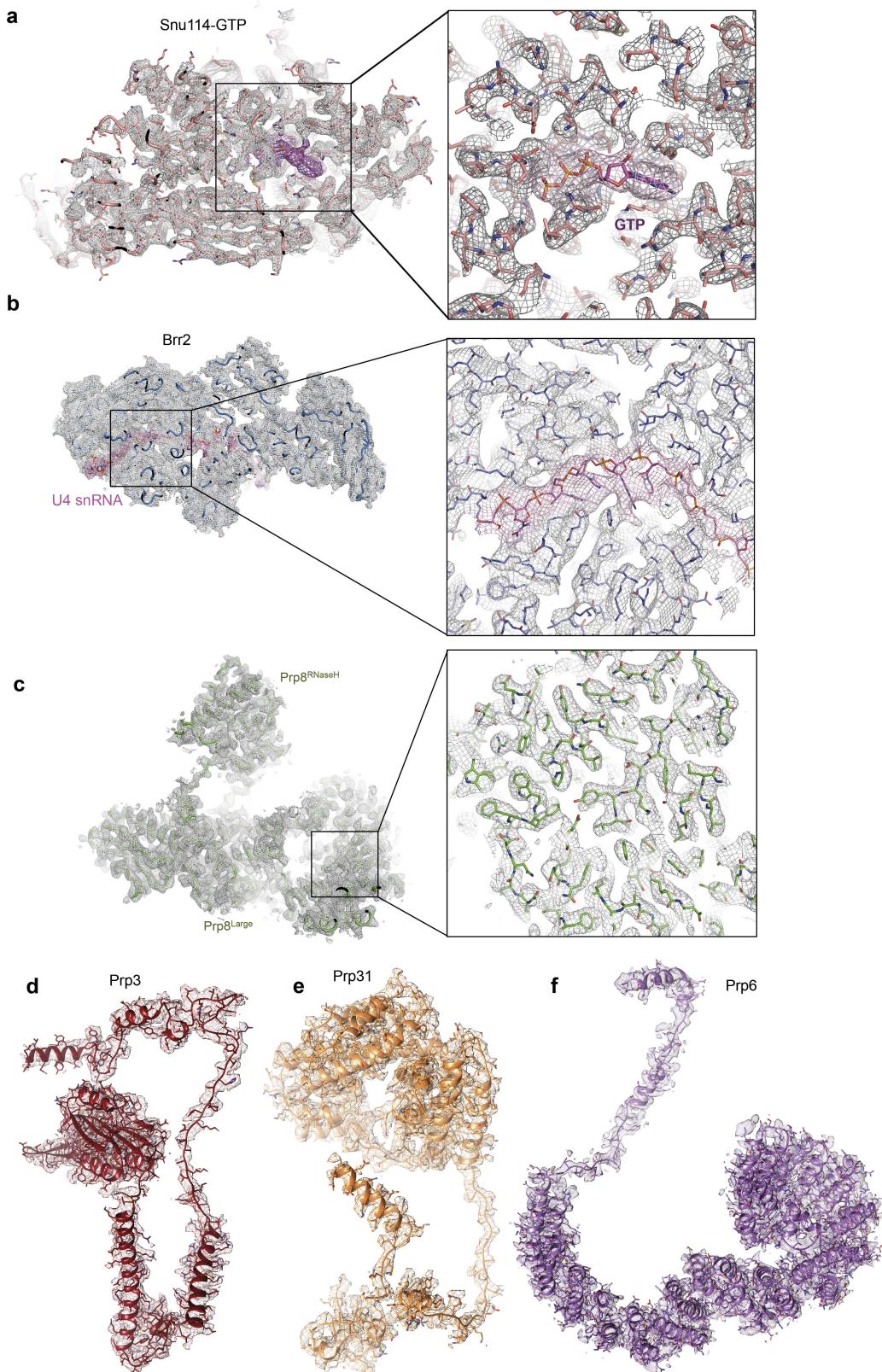


Extended Data Figure 1 | Image processing procedures. **a**, Representative micrograph. **b**, Representative 2D class averages obtained from reference-free 2D classification. **c**, Classification and refinement procedures used in this study.



Extended Data Figure 2 | Local and overall resolutions of tri-snRNP maps. Local resolution estimation by Resmap⁵⁵ of **a**, the overall 3.7 Å map and **b**, maps of the head, body and foot domains obtained from masked refinements with signal subtraction²³. **c**, Gold-standard FSC curves for the overall map and the maps of the head, body and foot domains obtained from masked refinements. Their resolutions are estimated at FSC = 0.143. **d–g**, FSC curves of model versus map and cross-validation

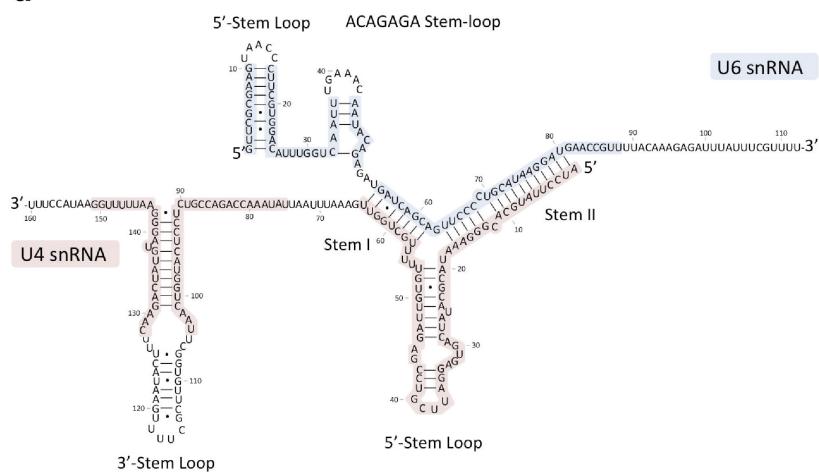
of model refinement by half-maps for the body, foot, head and overall maps, respectively. The red curves show FSC between the atomic model and the half-map it was refined against (half1) and the blue curves show FSC between the atomic model and the other half-map (half2) it was not refined against. The black curves show FSC between the atomic model and the sum map which the model was refined against.



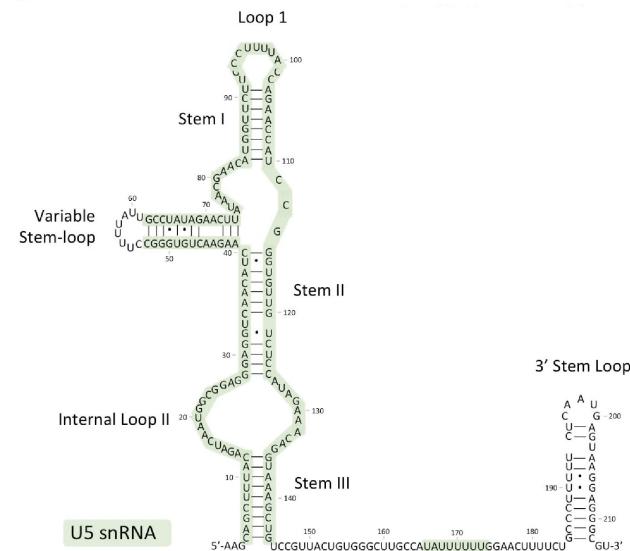
Extended Data Figure 3 | Representative EM density for different components of the map. **a**, Snu114 in the Foot domain with a bound GTP (magenta). The inset shows the GTP-binding pocket. **b**, Brr2 in the head domain with a bound single-stranded region of U4 snRNA. The inset

shows the density in the RNA binding tunnel. **c**, Density for Prp8 large and RNase-like domains. The inset shows the density in the core of Prp8. **d–f**, Prp3, Prp31 and Prp6 densities, respectively, with extended polypeptides.

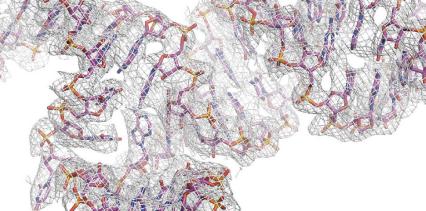
a



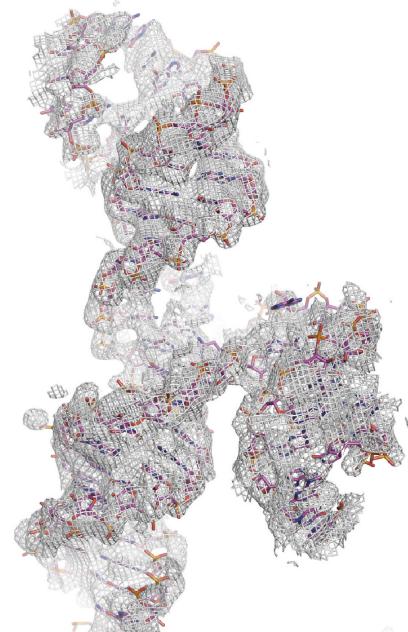
C



b

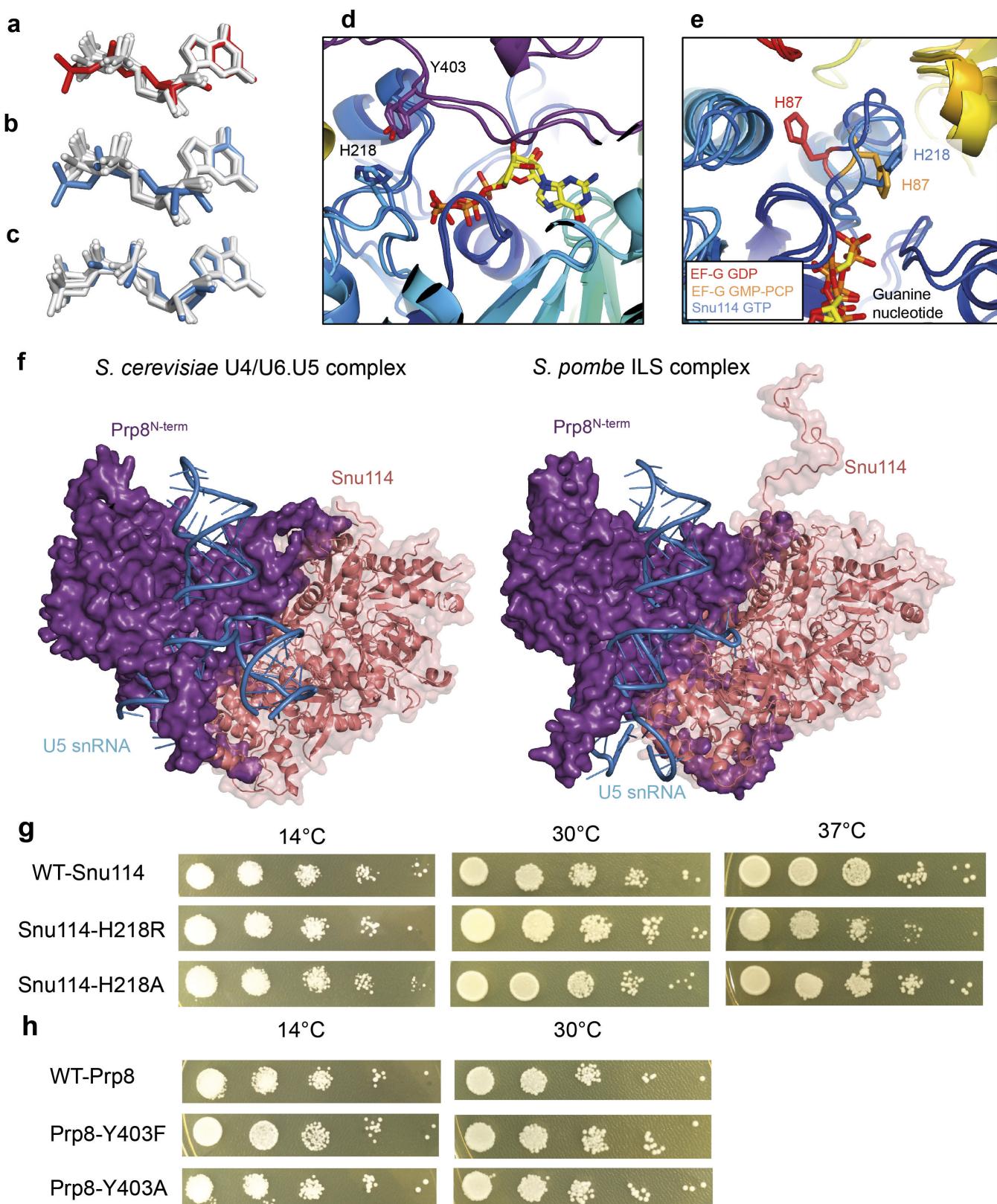
A 3D ribbon diagram of a protein-DNA complex. The DNA is represented by a series of interconnected grey and orange sticks forming a double helix. The protein structure is shown as a complex network of grey, orange, and purple sticks, forming a large, irregularly shaped cavity that encircles the DNA. A white arrow points to a specific region of the DNA double helix.

d



Extended Data Figure 4 | Secondary structure of the snRNAs in tri-snRNP. a, U4/U6 snRNA; c, U5 snRNA. The coloured nucleotides with red, green and blue background were built *de novo* into our EM density.

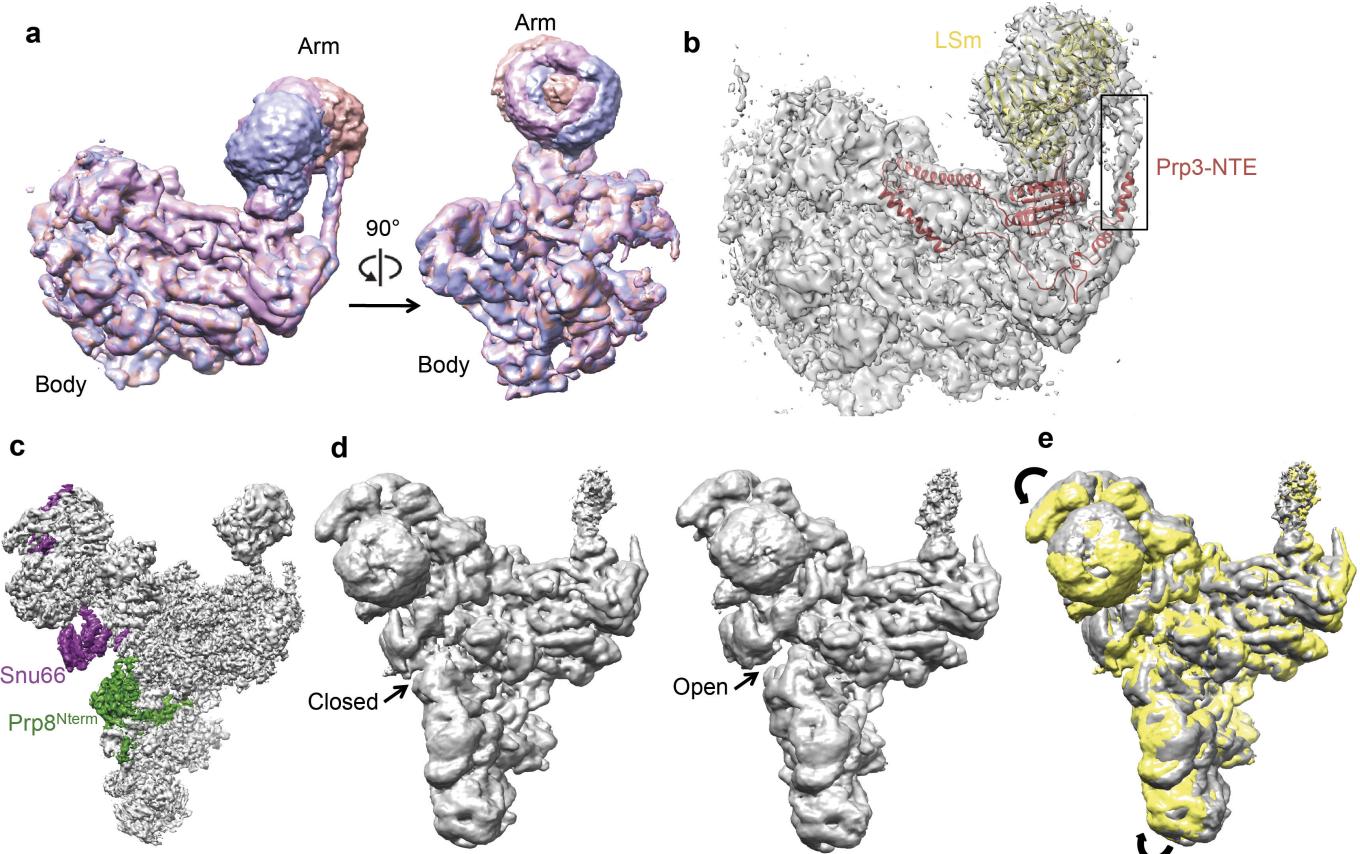
The region near the ACAGAGA sequence of U6 snRNA forms a stem-loop that was not predicted previously. **b, d**, Representative EM density for U4/U6 snRNA duplex and U5 snRNA, respectively.



Extended Data Figure 5 | See next page for figure caption.

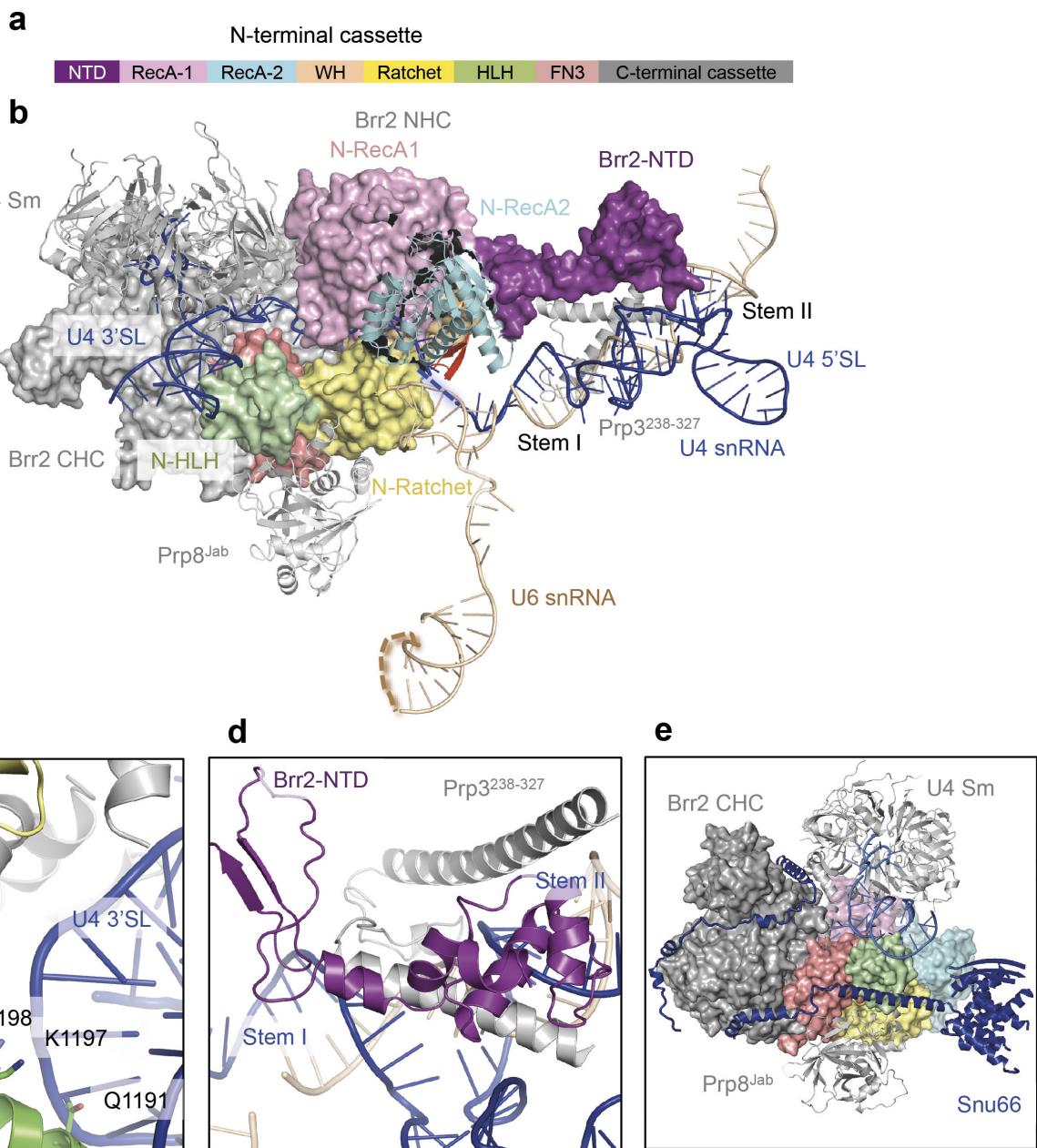
Extended Data Figure 5 | Interactions of Snu114 with guanine nucleotides and the N-terminal domain of Prp8 in the *S. cerevisiae* U4/U6.U5 tri-snRNP and *S. pombe* ILS complexes. **a**, Conformation of the Snu114(Cwf10)-bound GDP refined in the *S. pombe* ILS spliceosomal complex^{19,20} (red, PDB 3JB9), was overlaid on GDPs found in other guanine-nucleotide binding proteins (grey, PDB coordinates: 1DAR, 2E1R, 2WRI, 1Z0I, 5CA8, 1XTQ, 4YLG, 1SF8, 5BXQ). **b**, Guanine nucleotide refined as GDP in Snu114 of the *S. cerevisiae* U4/U6.U5 tri-snRNP (blue) is overlaid on GDPs found in the PDB coordinates as in **a**. **c**, Conformation of guanine nucleotide refined as GTP in Snu114 of the *S. cerevisiae* U4/U6.U5 tri-snRNP (blue) agrees well with GTP or GTP analogues in other guanine-nucleotide binding proteins (PDB code: 2BV3, 2DY1, 2J7K, 4YW9, 1ASO, 1LF0 (grey)). **d**, Superposition of the active site of

Snu114-GTP and Cwf10-GDP. **e**, Superposition of the GDP-bound EF-G (2WRI), GMP-PCP bound EF-G (4JUW) and Snu114 (*S. cerevisiae* tri-snRNP) active sites. His218 (His87 in EF-G) positions water molecule crucial for GTP hydrolysis. **f**, Comparison of Prp8^{N-term} domain, Snu114 and U5 snRNA in the *S. cerevisiae* U4/U6.U5 complex and *S. pombe* ILS complex. **g**, Growth of serial dilutions of yeast strains carrying wild-type Snu114, His218Arg or His218Ala Snu114 mutants at different temperatures. Cells were spotted on YPD plates and grown at 14 °C for 10 days, 30 °C and 37 °C for 2 days. **h**, Growth of serial dilutions of yeast strains carrying wild-type Prp8, Tyr403Phe and Tyr403Ala mutants. Cells were spotted on YPD plates and grown at 14 °C for 9 days, 30 °C for 3 days. This yeast strain does not survive at 37 °C and thus is not shown.



Extended Data Figure 6 | Conformational flexibility of tri-snRNP observed by classification. **a**, Different conformations of the arm domain demonstrated by the unsharpened maps of the three major classes (purple, magenta and red) obtained from masked classification of the arm domain alone followed by masked refinement with the body and arm domains. The body domain was included in the refinement because the arm domain is too small for accurate alignments. **b**, The sharpened map of one of the three classes with Prp3 and LSm models shown. In the improved domain maps for the arm domain, extra density for the N-terminal helix of Prp3

could be observed to extend to the LSm proteins. **c**, The sharpened map of the tri-snRNP and the locations of Snu66 and Prp8. **d**, The open and closed conformations of the head and foot domains of the tri-snRNP observed by global classification. The unsharpened maps for the two major classes obtained from global classification with finer angular sampling (1.8°) followed by 3D auto-refinement are shown. The open and closed states are indicated. **e**, Superposition of the unsharpened maps of the open (grey) and closed (yellow) states shown in **d**. The arrows indicate the rotations of the head and foot domains.



Extended Data Figure 7 | Brr2 helicase and its U4/U6 snRNA substrate.
a, Domain structure of Brr2 helicase comprising the N-terminal domain and two helicase cassettes. Individual domains of N-terminal helicase cassette (NHC) are colour-coded. **b**, Extensive interactions of Brr2 with U4/U6 snRNA and Prp3. The single-stranded region of U4 snRNA

extending from stem I enters the active site near the β -finger (red). **c**, 3' stem of U4 snRNA interacts with the HLH domain of NHC. **d**, The N-terminal domain (NTD) of Brr2 interacts with a long helix of Prp3 and inserts a loop into U4/U6 Stem II. **e**, Snu66 has a long extended region that wraps around both helicase cassettes of Brr2.

Extended Data Table 1 | Summary of model building of tri-snRNP components

	protein	total residues	M.W.	Modeled	Chain name	Local map	Human/ <i>S. pombe</i> names
U5 snRNP	Prp8	2413	279,299	110-2401	A	108-735: Foot 751-2104: Body 2147-2401: Head	220K/Spp42
	Brr2	2163	246,125	364-2163	B	363-433: Body 439-2163: Head	200K/Brr2
	Snu114	1008	114,025	102-989	C	Foot	116K/Cwf10
	Dib1	143	16,774	2-137	D	Body	15K/Dim1
	SmB	196	22,403	4-102	b		SmB/SmB
	SmD3	110	11,229	1-109	d		SmD3/SmD3
	SmD1	146	16,288	15-108	h		SmD1/SmD1
	SmD2	110	12,856	4-85	i	Foot	SmD2/SmD2
	SmE	94	10,373	4-92	e		SmE/SmE
	SmF	96	9,659	12-83	f		SmF/SmF
	SmG	77	8,479	2-76	g		SmG/SmG
U4/U6 snRNP	U5 snRNA-L	214	68,847	4-173	U	4-173: Foot 88-107: Body	
	Snu13	126	13,570	3-126	K	Body	15.5K/Snu13
	Prp31	494	56,305	43-457	F	Body	61K/Prp31
	Prp3	469	55,877	150-467	G	Body	90K/Prp3
	Prp4	465	52,425	109-465	H	Body	60K/Rna4
	SmB	196	22,403	4-102	k		SmB/SmB
	SmD1	146	16,288	1-118	l		SmD1/SmD1
	SmD2	110	12,856	15-108	m		SmD2/SmD2
	SmD3	110	11,229	4-85	n	Head	SmD3/SmD3
	SmE	94	10,373	10-92	p		SmE/SmE
tri-snRNP specific	SmF	96	9,659	12-83	q		SmF/SmF
	SmG	77	8,479	2-76	r		SmG/SmG
	Lsm2	95	11,164	1-90	2		Lsm2/Lsm2
	Lsm3	89	10,020	3-79	3		Lsm3/Lsm3
	Lsm4	172	20,304	1-90	4		Lsm4/Lsm4
	Lsm5	93	10,415	4-84	5	Arm	Lsm5/Lsm5
	Lsm6	86	9,396	11-84	6		Lsm6/Lsm6
	Lsm7	115	13,010	26-105	7		Lsm7/Lsm7
	Lsm8	109	12,385	1-67	8		Lsm8/Lsm8
	U4 snRNA	160	51,390	1-152	V	1-67: Body 73-152: Head	
tri-snRNP specific	U6 snRNA	112	36,088	1-112	W	1-26: Foot 26-88: Body 108-112: Arm	
	Prp6	899	104,234	155-898	J	Body	102K/Prp1
	Snu66	587	66,426	5-560 (poly-Ala)	E	Head	110K/Snu66
	Prp38	242	27,957	Not modeled			hPrp38/Prp38
	Snu23	194	22,682	Not modeled			hSnu23/Snu23
tri-snRNP specific	Spp381	291	33,764	Not modeled			

Extended Data Table 2 | Refinement, model statistics and structure/map depositions

a. Statistics of tri-snRNP structure determination**Data collection**

EM Titan Krios 300kV, K2 Gatan Summit

Pixel size (Å) 1.43

Defocus range (μm) -0.5 to -3.5

Reconstruction (RELION)

Overall 1.13 1.15 1.73 2.42

Accuracy of translations (pixel) 0.65 0.67 0.89 1.28

Final resolution (Å) 3.7 3.6 3.7 4.2

Refinement (REFMAC)

Refinement weight 0.001 0.001 0.001 0.001

Resolution limit (Å) 3.6 3.6 3.6 3.6

Residue numbers 9325 3728 2186 2922

Fourier Shell Correlation 0.75 0.85 0.82 0.60

R-factor (%) 29.7 27.8 28.7 31.5

Rms bond length (Å) 0.0078 0.0073 0.0073 0.011

Rms bond angle (°) 1.27 1.33 1.38 1.40

Ramachandran plot

Favoured 8066 (91.4%) 3266 (91.9%) 1810 (90.9%) 2531 (89.3%)

Allowed 615 (6.9%) 237 (6.7%) 135 (7.2%) 238 (9.3%)

Outliers 152 (1.7%) 50 (1.4%) 37 (1.9%) 39 (1.4%)

Validation by MolprobityGeometry score (percentile) 2.52 (98th) 2.41 (99th) 2.79 (95th) 2.62 (97th)Clash score (percentile) 7.48 (97th) 6.78 (100th) 11.4 (97th) 6.82 (100th)

Good rotamer (%) 94.8 95.7 93.5 93.2

b. Deposited maps and associated coordinate files

Maps	EMDB code	Associated PDB ID
Overall map	EMD-8012	5GAN
Body map	EMD-8014	5GAP
Head map	EMD-8013	5GAO
Foot map	EMD-8011	5GAM
Global class 1 (closed state)	EMD-8007	
Global class 2 (open state)	EMD-8006	
Masked body/arm class 1	EMD-8008	
Masked body/arm class 2	EMD-8009	
Masked body/arm class 3	EMD-8010	

Super-catastrophic disruption of asteroids at small perihelion distances

Mikael Granvik^{1,2}, Alessandro Morbidelli³, Robert Jedicke⁴, Bryce Bolin^{3,4}, William F. Bottke⁵, Edward Beshore⁶, David Vokrouhlický⁷, Marco Delbò³ & Patrick Michel³

Most near-Earth objects came from the asteroid belt and drifted via non-gravitational thermal forces into resonant escape routes that, in turn, pushed them onto planet-crossing orbits^{1–3}. Models predict that numerous asteroids should be found on orbits that closely approach the Sun, but few have been seen. In addition, even though the near-Earth-object population in general is an even mix of low-albedo (less than ten per cent of incident radiation is reflected) and high-albedo (more than ten per cent of incident radiation is reflected) asteroids, the characterized asteroids near the Sun typically have high albedos⁴. Here we report a quantitative comparison of actual asteroid detections and a near-Earth-object model (which accounts for observational selection effects). We conclude that the deficit of low-albedo objects near the Sun arises from the super-catastrophic breakup (that is, almost complete disintegration) of a substantial fraction of asteroids when they achieve perihelion distances of a few tens of solar radii. The distance at which destruction occurs is greater for smaller asteroids, and their temperatures during perihelion passages are too low for evaporation to explain their disappearance. Although both bright and dark (high- and low-albedo) asteroids eventually break up, we find that low-albedo asteroids are more likely to be destroyed farther from the Sun, which explains the apparent excess of high-albedo near-Earth objects and suggests that low-albedo asteroids break up more easily as a result of thermal effects.

Most near-Earth-object (NEO) discoveries during the past decade have been made by the Catalina Sky Survey (CSS), a combination of two distinct observatories with complementary capabilities. The 1.5-m Mt Lemmon telescope (code G96) typically detects faint NEOs close to the ecliptic, whereas the 0.8-m Catalina telescope (code 703) covers a larger area of the sky but focuses on brighter targets. From 2005 to 2012, CSS made 7,952 serendipitous detections of 3,632 distinct NEOs with absolute magnitudes H ranging from 17 to 25 during nights that had well established estimates for the detection efficiency. The detections were roughly equally shared between the two telescopes and the detected NEOs provide an extensive coverage of the parameter space (Supplementary Fig. 1).

We use the CSS NEO detections to constrain a model describing the true number of NEOs, $N(a, e, i, H)$, as a function of orbital parameters (semimajor axis a , eccentricity e and inclination i) and absolute magnitude H , a proxy for the physical size. Our new model of the NEO population is based on the methodology of ref. 3. First, we tracked the dynamical evolution of test asteroids from seven source regions or escape routes, s , in the main asteroid belt and nearby small-body reservoirs all the way into the inner Solar System (Supplementary Figs 2 and 3). Next, the orbital pathways followed by the bodies were assembled into source-specific steady-state orbital distributions, $R_s(a, e, i)$ (Supplementary Fig. 4). These functions were then multiplied by source-specific

parametric absolute-magnitude distributions, $N_s(H)$, and added together to produce $N(a, e, i, H)$. The model was then multiplied by the computed⁵ observational selection effects of CSS, $B(a, e, i, H)$ (Supplementary Fig. 5), thus obtaining a biased NEO model $N(a, e, i, H) \times B(a, e, i, H)$. The free parameters describing the $N_s(H)$ functions were determined by best-fitting the biased NEO model to the distribution of thousands of NEO detections from the CSS, $n(a, e, i, H)$, thus yielding a new and improved NEO population model. These computations are described in greater detail in Supplementary Information.

The observed a , e , i and H distributions generally agree with the biased NEO model (Supplementary Fig. 6). The $q = a(1 - e)$ distribution reveals a systematic offset in that the model predicts too many NEOs with small q and too few with large q (Fig. 1). If we assume that the overprediction at $q \leq 0.6$ astronomical units (AU) is the real source for the discrepancy, then the underprediction at large q is a feature resulting from how we fit the model to the data: the absolute number of NEO detections is one of the constraints and therefore the method compensates for an overprediction at small q with an underprediction at large q . The discrepancy in the q distribution has not been noticed in the past because previous models were calibrated with much smaller samples of NEOs and the discrepancy was not statistically significant⁶.

A possible explanation for the discrepancy is that the combination of orbital steady-state distributions $R_s(a, e, i)$ is not flexible enough to allow for a good fit at small q . To test this explanation we divided the test asteroids into a larger number N_s of source regions and re-did the analysis. It turns out that even a model with $N_s = 24$ (Supplementary Fig. 7) is unable to match the observed q distribution (Supplementary Fig. 8). To rule out systematic problems with our orbital steady-state distributions we also tried orbital steady-state distributions computed by others using different starting conditions and integration parameters⁷. The alternative orbital distributions resulted in an even worse fit to the observed data, at least partly explained by the smaller number of source regions considered (Supplementary Fig. 9), and do not solve the discrepancy in q (Supplementary Fig. 8).

To validate the adopted methodology and, in particular, the bias correction, we carried out two tests which made use of the fact that CSS is composed of two surveys, G96 and 703, with partly complementary capabilities. The first test was to fit separately to G96 and to 703 to identify problems in either one of the surveys or their estimated selection effects. The discrepancy in q is present in both cases (Supplementary Fig. 8) and we conclude that the discrepancy in q is not specific to the chosen survey. The second test was a cross-check of the results: we estimated model parameters by fitting just the data from 703 (G96) in the limited range $0.7 \text{ AU} < q < 1.3 \text{ AU}$, used the resulting model and estimates for selection effects to predict the absolute number of detections for G96 (703) in the same q range, and compared the prediction to the data. The results show that the adopted methodology and bias corrections are sound

¹Department of Physics, PO Box 64, 00014 University of Helsinki, Finland. ²Finnish Geospatial Research Institute, PO Box 15, 02430 Masala, Finland. ³Observatoire de la Côte d'Azur, Boulevard de l'Observatoire, F 06304 Nice Cedex 4, France. ⁴Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, Hawaii 96822, USA. ⁵Southwest Research Institute, 1050 Walnut Street, Suite 300, Boulder, Colorado 80302, USA. ⁶University of Arizona, 933 North Cherry Avenue, Tucson, Arizona 85721-0065, USA. ⁷Institute of Astronomy, Charles University, V Holešovičkách 2, CZ 18000 Prague 8, Czech Republic.

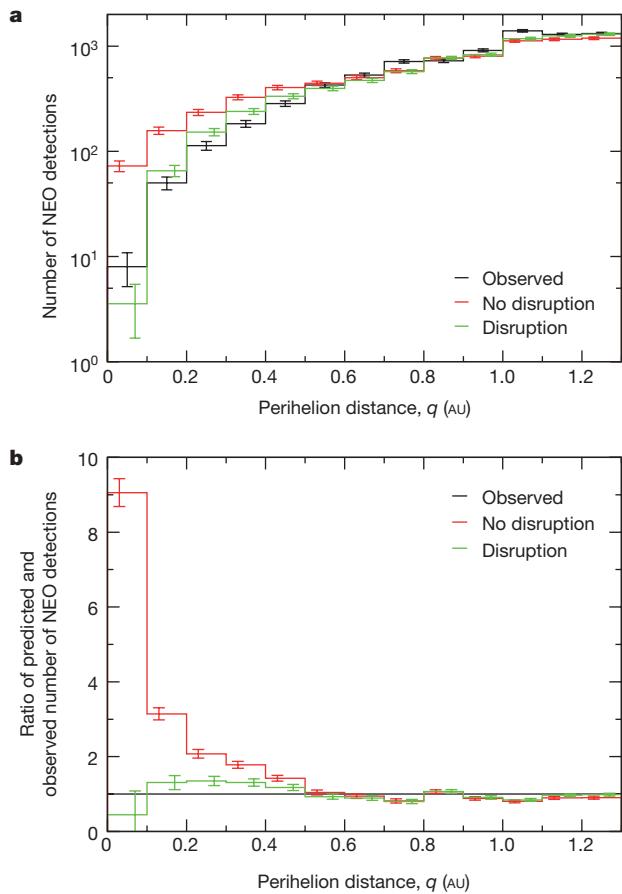


Figure 1 | Observed and predicted perihelion distances for NEOs detected by the CSS during 2005–2012. The observed perihelion distance distribution (black) compared with the biased model predictions with (green) and without (red), assuming a disruption at $\bar{q}_* = 0.076$ AU. **a**, The observed and predicted number of NEO detections. **b**, The ratio between the observed and predicted number of NEO detections. The model without disruption shows a systematic overprediction at small q , whereas assuming a disruption breaks the trend and reproduces the observed q distribution. The s.e.m. error bars are computed assuming Poisson statistics.

and that the simultaneous use of complementary data sets leads to improved and more accurate models (Supplementary Fig. 10). Extrapolations to $q < 0.7$ AU systematically show that CSS should have discovered more NEOs on orbits with small q if such objects existed (Supplementary Fig. 8). We therefore conclude that some physical mechanism must be reducing the number of NEOs at small q .

We propose that when NEOs reach some critical perihelion distance, q_* , they catastrophically disrupt. To test the proposed mechanism we constructed alternative orbital steady-state distributions by considering the dynamical evolution of test asteroids only until their q becomes smaller than a pre-defined value q_* . The observed q distribution is best reproduced with $\bar{q}_* = (0.0760 \pm 0.0025)$ AU $\approx (16 \pm 0.5)R_\odot$ (Fig. 1). Also the observed (a, e, i, H) distributions are accurately reproduced (Supplementary Fig. 11). The best-fit model predicts that there are $(7.32 \pm 1.33) \times 10^5$ NEOs with $17 < H < 25$ and $1,008 \pm 45$ NEOs with $H < 17.75$. Both numbers agree with the most recent estimates⁸ (Supplementary Fig. 12). The best-fit model also reproduces the observed relative fractions of Amor ($1.017 \text{ AU} < q \leq 1.3 \text{ AU}$), Apollo ($a \geq 1 \text{ AU}$ and $q \leq 1.017 \text{ AU}$) and Aten ($a < 1 \text{ AU}$ and aphelion distance $Q > 0.983 \text{ AU}$) asteroids with $17 \leq H < 17.5$ —the observed fractions are 47%, 50% and 3%, whereas the model predicts $43 \pm 5\%$, $53 \pm 5\%$ and $3.5 \pm 0.6\%$, respectively. See the Supplementary Video for an animation of how the orbit distribution changes with H .

To assess the effect of size on \bar{q}_* , we divided the NEOs detected by CSS into three different groups as a function of H , and fitted \bar{q}_* separately to each of these groups. The result shows a clear trend of increasing \bar{q}_* with increasing H (Fig. 2), that is, an inverse correlation between \bar{q}_* and physical size. A direct consequence of the inverse correlation is that a kilometre-scale asteroid has to disrupt into fragments smaller than a few tens of metres in a single event or through a disruption cascade, depending on the disruption mechanism. The disruption distances are too large to be explained by tidal effects and evaporation⁹. While the average surface temperature of the sunlit hemisphere on mid-sized NEOs may surpass 900 K, the resulting diurnal heat waves will penetrate¹⁰ only to depths of some tens of centimetres. Silicates immediately below the surface layer will therefore remain solid. This reveals that the actual disruption mechanism, although clearly related to temperature, is not trivial. A possibility is that rocks break into small grains by thermal cracking¹¹ and the grains are then blown away from the asteroid by radiation pressure¹². Another possibility is that the

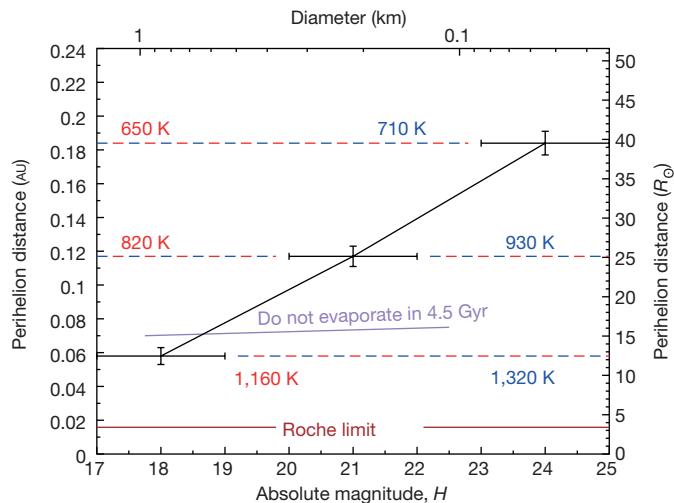


Figure 2 | Average disruption distance as a function of absolute magnitude and an asteroid's physical size. We split the NEO detections by CSS into three different H groups, with the H range shown by the horizontal error bars ($17 < H < 19$ contains 3,326 detections, $20 < H < 22$ contains 1,669 detections, and $23 < H < 25$ contains 913 detections). The average dynamical lifetime of NEOs is a few million years and asteroids on circular orbits with $r = \bar{q}_*$ above the purple line will not have evaporated⁹ after 4.5 billion years, the age of the Solar System. Evaporation can thus not explain the disappearance of small and mid-sized NEOs and, given the timescales, it is also an unlikely disruption mechanism for large NEOs. The brown horizontal line marks the Sun's Roche limit for a hypothetical fluid comet with a density of 0.5 g cm^{-3} , and serves as an approximate upper limit for tidal disruption of asteroids and comets. The red dashed lines correspond to the equilibrium temperature, $T_{\text{eq}} = [(1 - A)L_0 / (16\pi\sigma r^2)]^{0.25}$, at perihelion, $r = q$, when assuming a Bond albedo of $A = 0.07$ and an infrared emissivity of $\varepsilon = 0.9$. $L_0 = 3.827 \times 10^{26} \text{ W}$ is the solar luminosity and $\sigma = 5.6704 \times 10^8 \text{ W m}^{-2} \text{ K}^{-4}$ is the Stefan–Boltzmann constant. The blue dashed lines correspond to the simple estimate of the temperature average over a sunlit hemisphere $T_{\text{av}} = 4T_{\text{ss}}/5$ with the subsolar temperature $T_{\text{ss}} = \sqrt{2}T_{\text{eq}}$. The true average surface temperature will lie somewhere between T_{eq} and T_{av} , because T_{eq} does not allow for local temperature variations and T_{av} does not account for conduction and sublimation. The conversion between H and diameter assumes a geometric albedo of 0.15. The detection-weighted average \bar{q}_* of the three groups is 0.094 ± 0.010 AU, which is about $24 \pm 17\%$ larger than the value obtained by fitting all groups simultaneously. The difference is a systematic error resulting from averaging over H . The line connecting the three groups emphasizes the linear (nonlinear) relation between the \bar{q}_* and H (diameter). The s.d. error bars on \bar{q}_* were estimated by generating 50 random representations of the best-fit model and re-fitting for \bar{q}_* . We required that the solutions for \bar{q}_* must reproduce the observed q distribution: that is, all \bar{q}_* that were substantially larger than the smallest observed q were discarded.

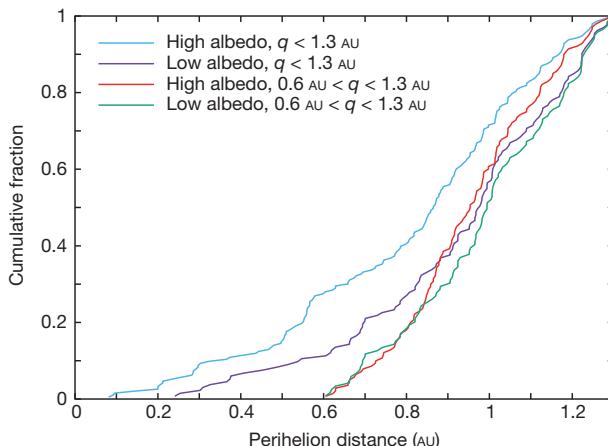


Figure 3 | Empirical distribution functions in perihelion distance for low-albedo and high-albedo NEOs detected by WISE. The WISE data set is biased against high-albedo NEOs: for a given absolute magnitude H , the higher the geometric albedo, the fainter the infrared apparent magnitude $W3$. The WISE detection efficiency is close to 100% up to $W3 = 9.5$, drops to 50% at $W3 = 10$ and is close to zero beyond²² $W3 = 10.5$. This means that asteroids with large enough albedos would not have been observed by WISE. To correct for the WISE albedo bias we assume that its limiting magnitude is $W3_{\text{lim}} = 10$. Then, for each NEO with a WISE-determined albedo we identify all its WISE-reported observations at different epochs, select the smallest apparent magnitude $W3_{\text{min}}$ of all reported values, and reject that NEO from consideration if $W3_{\text{min}} > W3_{\text{lim}}$. This reduces the initial number of 394 NEOs to 326 with $W3_{\text{lim}} = 10$ and $H > 15$, the latter requirement ensuring that the albedos correspond to relatively small NEOs. The Anderson–Darling test applied to empirical distribution functions in q shows that it is extremely unlikely that the 133 low-albedo and 193 high-albedo NEOs have a common parent distribution ($p \approx 0.0003$) when accounting for $q < 1.3$ AU. A common origin for the q distributions becomes reasonable when limiting the analysis to NEOs with $0.6 \text{ AU} < q < 1.3$ AU only, that is, 119 low-albedo and 140 high-albedo NEOs ($p \approx 0.12$).

anisotropic emission of thermal photons¹³ or the scattering of sublimating gas molecules¹⁴ cause the asteroids to spin faster, to the point when gravity and cohesive forces can no longer keep them intact. A third possibility is that all asteroids contain volatile elements that, when sublimating at a moderate temperature, exert enough pressure on the body to blow it up.

To gain insight into the process leading to asteroid disruption, we investigated whether asteroids with different surface properties behave differently. For this purpose, we compared the q distributions of low-albedo and high-albedo NEOs detected by the Wide-field Infrared Survey Explorer (WISE) mission. The Anderson–Darling test¹⁵ shows that the probability that these two samples come from a common parent distribution is only 0.03%, whereas a reasonable agreement is found when limiting the analysis to NEOs with $q > 0.6$ AU (Fig. 3). This result agrees with the results of an independent analysis⁴ of WISE data, which showed that the observed Aten asteroids have, on average, higher albedos than Apollo and Amor asteroids. This can be explained if low-albedo NEOs disrupt, on average, farther from the Sun than high-albedo asteroids of comparable size, implying that they have different physical properties that make them more vulnerable to strong solar irradiation. The fact that the q distribution of low-albedo NEOs appears to be steeper than that of high-albedo NEOs at $q < 0.6$ AU supports this conclusion: the larger q^* is, the steeper is the resulting q distribution (Supplementary Fig. 13). A larger average disruption distance may be due to a higher volatile budget in low-albedo asteroids, as suggested by the composition of the most primitive carbonaceous meteorites (usually expected to be related to these bodies) and by the quite common presence of hydration bands in their spectra. Thermal cracking is also more efficient for carbonaceous meteorites than for ordinary chondrites (the meteorites associated with high-albedo asteroids). We also

note that darker NEOs experience somewhat greater heating and may therefore be more prone to thermally driven disruption.

In our model we assume that an instantaneous disruption takes place when $q \leq \bar{q}_*$ for an NEO. We predict that NEO disruptions must take place in less than 250 years, the timescale used to record our model data. Our results are consistent with the (q, H) distribution of known asteroids. Asteroid (394130) 2006 HY₅₁ has the smallest perihelion distance, $q \approx 0.081$ AU $\approx 17.4R_{\odot}$, among known NEOs with reliable estimates for the absolute magnitude. Its absolute magnitude, $H = 17.2$, is in agreement with our assessment of the average disruption distance (Fig. 2). Our results are also in agreement with observations of a slow erosion of the asteroid (3200) Phaethon¹⁶, which is too large to disrupt catastrophically on its current orbit.

The recorded inclination distribution of test asteroids at the disruption epoch overlaps with the observed inclination distribution of $q < 0.184$ AU meteor showers identified in data obtained by the Canadian Meteor Orbit Radar^{17,18} (CMOR; Supplementary Fig. 14). While covering the same range, the latter distribution is skewed towards larger values, which can be understood considering that radar is more sensitive to high-speed meteors and hence the orbital distribution is biased against low-inclination orbits. Super-catastrophic disruptions are consistent with the fact that parent bodies have yet to be detected for most meteor streams with small q and small i identified in CMOR data. Given the inverse correlation between \bar{q}_* and asteroid diameter we predict that the average total mass of meteor streams lacking obvious parent bodies should diminish as a function of q as long as $q \leq 0.2$ AU and $i \leq 40^\circ$.

In the future, a detailed understanding of the circumstances leading to disruption of asteroids at different q values may offer insight into their bulk composition as well as internal structure. In particular, a quantitative assessment of the volatile content for NEOs, and hence of their siblings in the main asteroid belt, by mapping out the disruption probability as a function of q and source region, would complement current approaches, which usually rely on extrapolation from surface properties¹⁹ and the detection of comet-like activity^{20,21}.

Received 13 July; accepted 7 December 2015.

1. Morbidelli, A. & Vokrouhlický, D. The Yarkovsky-driven origin of near Earth asteroids. *Icarus* **163**, 120–134 (2003).
2. Morbidelli, A., Bottke, W. F. Jr, Froeschlé, C. & Michel, P. Origin and evolution of near-Earth objects. *Asteroids III*, 409–422 (2002).
3. Bottke, W. F., Jedicke, R., Morbidelli, A., Petit, J.-M. & Gladman, B. Understanding the distribution of near-Earth asteroids. *Science* **288**, 2190–2194 (2000).
4. Mainzer, A. *et al.* Characterizing subpopulations within the near-Earth objects with NEOWISE: preliminary results. *Astrophys. J.* **752**, 110 (2012).
5. Jedicke, R., Bolin, B., Granvik, M. & Beshore, E. A fast method for quantifying observational selection effects in asteroid surveys. *Icarus* **266**, 173–188 (2016).
6. Bottke, W. F. *et al.* Debiased orbital and absolute magnitude distribution of the near-Earth objects. *Icarus* **156**, 399–433 (2002).
7. Greenstreet, S., Ngo, H. & Gladman, B. The orbital distribution of near-Earth objects inside Earth's orbit. *Icarus* **217**, 355–366 (2012).
8. Harris, A. W. & D'Abramo, G. The population of near-Earth asteroids. *Icarus* **257**, 302–312 (2015).
9. Campins, H., Davis, D. R., Weidenschilling, S. J. & Magee, M. in *Completing the Inventory of the Solar System* (eds Rettig, T. & Hahn, J. M.) *Astron. Soc. Pacif. Conf. Ser.* **107**, 85–96 (1996).
10. Lewis, J. S. *Physics and Chemistry of the Solar System* 257–259 (Elsevier Academic Press, 2004).
11. Delbo, M. *et al.* Thermal fatigue as the origin of regolith on small asteroids. *Nature* **508**, 233–236 (2014).
12. Jewitt, D. The active asteroids. *Astron. J.* **143**, 66 (2012).
13. Vokrouhlický, D., Bottke, W. F., Chesley, S. R., Scheeres, D. J. & Statler, T. S. The Yarkovsky and YORP effects. *Asteroids IV*, 509–531 (2015).
14. Steckloff, J. K. & Jacobson, S. A. The formation of striae within cometary dust tails by a sublimation-driven YORP-like effect. *Icarus* **264**, 160–171 (2016).
15. Feigelson, E. D. & Jogesh Babu, G. *Modern Statistical Methods for Astronomy* 105–117 (Cambridge Univ. Press, 2012).
16. Jewitt, D., Li, J. & Agarwal, J. The dust tail of asteroid (3200) Phaethon. *Astrophys. J.* **771**, L36 (2013).
17. Brown, P., Weryk, R. J., Wong, D. K. & Jones, J. A meteoroid stream survey using the Canadian Meteor Orbit Radar. I. Methodology and radiant catalogue. *Icarus* **195**, 317–339 (2008).

18. Brown, P., Wong, D. K., Weryk, R. J. & Wiegert, P. A meteoroid stream survey using the Canadian Meteor Orbit Radar. II: identification of minor showers using a 3D wavelet transform. *Icarus* **207**, 66–81 (2010).
19. Rivkin, A. S., Howell, E. S., Vilas, F. & Lebofsky, L. A. Hydrated minerals on asteroids: the astronomical record. *Asteroids III*, 235–253 (2002).
20. Hsieh, H. H. & Jewitt, D. A population of comets in the main asteroid belt. *Science* **312**, 561–563 (2006).
21. Küppers, M. *et al.* Localized sources of water vapour on the dwarf planet (1) Ceres. *Nature* **505**, 525–527 (2014).
22. Mainzer, A. *et al.* NEOWISE observations of near-Earth objects: preliminary results. *Astrophys. J.* **743**, 156 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements Discussions with D. Nesvorný, K. Tsiganis and S. Jacobson as well as a review by A. Harris helped improve the manuscript. M.G. was funded by grant 137853 from the Academy of Finland and D.V. by grant GA13-01308S of the Czech Science Foundation. W.F.B. thanks NASA's Near Earth Object Observation programme for supporting his

work in this project. We acknowledge support by ESA via contract AO/1-7015/11/NL/LvH (Synthetic Generation of a NEO Population). CSC-IT Centre for Science Ltd, Finland, the Finnish Grid Infrastructure and the mesocentre SIGAMM at the Observatoire de la Côte d'Azur provided computational resources.

Author Contributions M.G. conceived the disruption model. M.G., A.M. and W.F.B. carried out orbital integrations using code modified by D.V. and A.M. R.J., B.B., M.G. and E.B. estimated observational biases for CSS. M.G. developed the code for fitting model parameters and carried out the fits. M.D. provided expertise on the WISE albedo analysis. M.G. wrote the paper with input from all authors.

Author Information The code for orbital integrations, SWIFT, is available at <https://www.boulder.swri.edu/~hal/downloads>. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.G. (mgranvik@iki.fi).

Universal resilience patterns in complex networks

Jianxi Gao^{1*}, Baruch Barzel^{2*} & Albert-László Barabási^{1,3,4,5}

Resilience, a system's ability to adjust its activity to retain its basic functionality when errors, failures and environmental changes occur, is a defining property of many complex systems¹. Despite widespread consequences for human health², the economy³ and the environment⁴, events leading to loss of resilience—from cascading failures in technological systems⁵ to mass extinctions in ecological networks⁶—are rarely predictable and are often irreversible. These limitations are rooted in a theoretical gap: the current analytical framework of resilience is designed to treat low-dimensional models with a few interacting components⁷, and is unsuitable for multi-dimensional systems consisting of a large number of components that interact through a complex network. Here we bridge this theoretical gap by developing a set of analytical tools with which to identify the natural control and state parameters of a multi-dimensional complex system, helping us derive effective one-dimensional dynamics that accurately predict the system's resilience. The proposed analytical framework allows us systematically to separate the roles of the system's dynamics and topology, collapsing the behaviour of different networks onto a single universal resilience function. The analytical results unveil the network characteristics that can enhance or diminish resilience, offering ways to prevent the collapse of ecological, biological or economic systems, and guiding the design of technological systems resilient to both internal failures and environmental changes.

The traditional mathematical treatment of resilience used from ecology⁴ to engineering⁵ approximates the behaviour of a complex system with a one-dimensional (1D) nonlinear dynamic equation

$$\frac{dx}{dt} = f(\beta, x) \quad (1)$$

The functional form of $f(\beta, x)$ represents the system's dynamics, and the parameter β captures the changing environmental conditions. The system is assumed to be in one of the stable fixed points, x_0 , of equation (1), extracted from⁵

$$f(\beta, x_0) = 0 \quad (2)$$

$$\left. \frac{\partial f}{\partial x} \right|_{x=x_0} < 0 \quad (3)$$

where equation (2) provides the system's steady state and equation (3) guarantees its linear stability. The solution of equations (2) and (3) provides the resilience function $x(\beta)$, which represents the possible states of the system as a function of β (Fig. 1a–c). The shape of this function is uniquely determined by the functional form of $f(\beta, x)$. In contrast, the momentary state of the system is determined by the tunable parameter β . At some critical point β_c the resilience function may feature a bifurcation (Fig. 1a) or become non-analytic (Fig. 1b, c), indicating that

the system loses its resilience by undergoing a sudden transition to a different^{8,9}, often undesirable, fixed point of equation (1).

Although it is conceptually powerful, this analytic framework does not account for the exceptionally large number of variables that in reality control the state of a complex system. Indeed, real systems are composed of numerous components linked via a complex set of weighted, often directed, interactions^{10,11}, and controlled by not one microscopic parameter, but by a large family of parameters, such as the weights of all interactions. Hence, instead of a 1D function $f(\beta, x)$, characterized by a single parameter β , their state should be described by a network of coupled nonlinear equations that capture the interactions between the system's many components, and account for the complex interplay between the system's dynamics and changes in the underlying network topology^{6,12}. The resulting resilience function is therefore a multi-dimensional manifold over the complex parameter space characterizing the system (Fig. 1d–f), which, using the current tools, cannot be treated analytically. Here we overcome these longstanding limitations by developing a general network-based theoretical framework that allows us to explore and predict the multiple roots and dimensions of resilience, exposing crucial determinants of resilience loss in complex natural and man-made systems.

Consider a system consisting of N components (nodes) whose activities $\mathbf{x} = (x_1, \dots, x_N)^\top$ follow the coupled nonlinear equations^{12,13}

$$\frac{dx_i}{dt} = F(x_i) + \sum_{j=1}^N A_{ij} G(x_i, x_j) \quad (4)$$

The first term on the right-hand side of equation (4) describes the self-dynamics of each component, while the second term describes the interactions between component i and its interacting partners. The nonlinear functions $F(x_i)$ and $G(x_i, x_j)$ represent the dynamical laws that govern the system's components, while the weighted connectivity matrix A_{ij} captures the interactions between the nodes. With an appropriate choice of $F(x_i)$ and $G(x_i, x_j)$, equation (4) is used to model numerous systems known for their resilience, ranging from cellular¹⁴ to ecological^{15,16} and social systems¹⁷.

In analogy with the 1D system of equation (1), a transition from a desired to an undesired stable fixed point captures the loss of resilience in the multi-dimensional system of equation (4). The key difference is that in equation (4) resilience loss can be induced by changes in any of the N^2 parameters of the weighted network A_{ij} , each change capturing a different kind of perturbation (Fig. 1g). For instance, the extinction/introduction of species in an ecological system may correspond to the removal/addition of one or several nodes^{7,18}. Alternatively, the loss of an enzyme catalysing a reaction in a metabolic network¹⁹ may correspond to link removal. Finally, global environmental changes, such as increases in ocean acidity or global warming²⁰, may correspond to global shifts in the weights of A_{ij} .

We illustrate the challenges such multi-dimensional systems offer by exploring the mutualistic interactions among species in an ecological

¹Center for Complex Network Research, Department of Physics, Northeastern University, Boston, Massachusetts 02115, USA. ²Department of Mathematics, Bar-Ilan University, Ramat-Gan 52900, Israel. ³Center for Cancer Systems Biology, Dana-Farber Cancer Institute, Harvard University, Boston, Massachusetts 02215, USA. ⁴Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁵Center for Network Science, Central European University, Budapest 1051, Hungary.

*These authors contributed equally to this work.

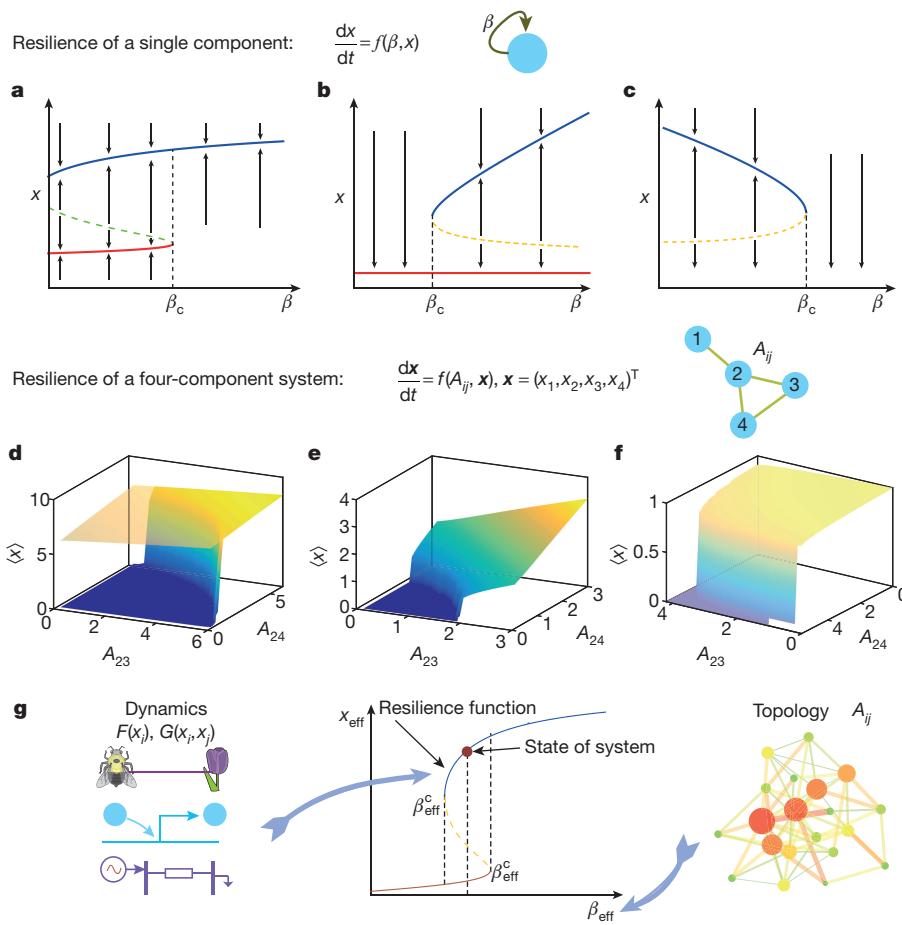


Figure 1 | Network resilience. **a–c**, In 1D systems resilience is captured by the resilience function $x(\beta)$, which describes the state(s) of the system as a function of the tunable parameter β . We illustrate three typical examples. **a**, The bifurcating resilience function. The system exhibits a single stable fixed point for $\beta > \beta_c$ (blue) and two (or more) stable fixed points, a desired (blue) and an undesired (red) for $\beta < \beta_c$. **b**, Resilience function with a first-order transition from the desired (blue) state to the undesired (red) state. **c**, Resilience function with a stable solution for $\beta < \beta_c$ and no solution above β_c , resulting in an uncontrolled divergent or chaotic behaviour. **d–f**, In a multi-dimensional system, the single parameter β is replaced by the complex weighted network A_{ij} , whose characteristics depend on both environmental conditions and the specific pairwise interaction strengths. Consequently, the resilience function, now capturing the behaviour of the vector state $x(A_{ij})$, is a multi-dimensional manifold, prohibiting analytical treatment. The 3D plots show the resilience plane for a four-node system, displaying $x(A_{23}, A_{24})$ for fixed A_{12} and A_{34} . The full description of an N -dimensional system requires an N^2 -dimensional plane, tracking the state of the system as a function of all A_{ij} . **g**, After applying our formalism, the multi-dimensional manifold shown in **d–f** collapses into a 1D resilience function in β -space (blue and red solid lines). The structure of this function, and hence its critical points β_{eff}^c (dashed lines) is fully determined by the system's dynamics $F(x_i)$ and $G(x_i, x_j)$: ecological, regulatory, power transmission and so on (left). The network topology A_{ij} (right) determines β_{eff} through equation (8), and hence the specific state of the system (brown dot) along the resilience function.

network. Here equation (4) tracks the abundance $x_i(t)$ of species i , following¹⁶

$$\frac{dx_i}{dt} = B_i + x_i \left(1 - \frac{x_i}{K_i}\right) \left(\frac{x_i}{C_i} - 1\right) + \sum_{j=1}^N A_{ij} \frac{x_i x_j}{D_i + E_i x_i + H_j x_j} \quad (5)$$

The first term on the right hand side of equation (5) accounts for the incoming migration of i at a rate B_i from neighbouring ecosystems. The second term describes logistic growth with the system carrying capacity²¹ K_i , and the Allee effect, according to which for low abundance ($x_i < C_i$) the system features negative growth²². The third term describes mutualistic interactions, captured by a response function that saturates for large x_i or x_j , indicating that j 's positive contribution to x_i is bounded. To construct A_{ij} we used symbiotic interactions, such as plant–pollinator relationships, collected for seven ecological systems²³ (Supplementary Table 1), describing networks ranging from $N = 10$ to $N = 1,429$ nodes (Fig. 2a, b). We numerically solved equation (5), and tested its resilience under three realistic perturbations (Fig. 2c): first, we randomly removed a fraction f_n of nodes, capturing plant extinctions; second, we removed a random fraction f_l of pollinators, thus perturbing the mutualistic link weights; and finally, we randomly rescaled all the weights A_{ij} , reducing their strength on average by a fraction f_w to mimic global environmental changes.

We find that for small perturbations the system maintains its resilience: its only stable fixed point is x^H , in which the average abundance $\langle x \rangle$ is high. However, when the perturbation exceeds a certain threshold a bifurcation occurs, resulting in two stable fixed points: the desired x^H and an undesired low-abundance state x^L (Fig. 2d). Under these conditions the system loses its resilience, potentially transitioning to the undesired x^L . The precise bifurcation point marking this loss of resilience is, however, highly unpredictable. For instance,

Net1 displays a different resilience pattern for node removal (Fig. 2d), link removal (Fig. 2e) or global weight changes (Fig. 2f), indicating that different forms of perturbations lead to different outcomes within the same system. Such diversity is also observed for similar perturbations across different systems. For example, while we need to remove at least 35% of the pollinators for Net1 to lose its resilience (Fig. 2e), Net5 remains resilient even after 80% of its pollinators are deleted (Fig. 2h). Finally, even the sequence in which the perturbation is applied makes an important difference: Net1 can lose its resilience anywhere between the removal of 30% to 70% of its nodes (Fig. 2d), depending on the specific trial.

Altogether, we analysed fourteen experimentally mapped networks (Supplementary Table 1), finding that the resilience transition depends on the network topology, the form and the nature of the perturbation applied and the specific realization (Fig. 2d–l, and Supplementary Figs 2–4), exposing severe limits to our ability to predict the network resilience. We now show that this seemingly unpredictable behaviour can be systematically treated by focusing on the system's natural state and control variables. In a network environment, the state of each node is affected by the state of its immediate neighbours. Therefore, we characterize the effective state of the system using the average nearest-neighbour activity

$$x_{\text{eff}} = \frac{\mathbf{I}^\top A \mathbf{x}}{\mathbf{I}^\top A \mathbf{I}} = \frac{\langle s^{\text{out}} x \rangle}{\langle s \rangle} \quad (6)$$

where $s^{\text{out}} = (s_1^{\text{out}}, \dots, s_N^{\text{out}})^\top$ is the vector of outgoing weighted degrees and $s^{\text{in}} = (s_1^{\text{in}}, \dots, s_N^{\text{in}})^\top$ is the vector of incoming weighted degrees, $\langle s^{\text{out}} x \rangle = (1/N) \sum_{i=1}^N s_i^{\text{out}} x_i$, $\langle s \rangle = \langle s^{\text{in}} \rangle = \langle s^{\text{out}} \rangle$ is the average weighted degree, and \mathbf{I} is the unit vector $\mathbf{I} = (1, \dots, 1)^\top$. As we show in Supplementary Information section I, if A_{ij} has not much degree correlation, the variable x_{eff} in equation (6) allows us to reduce the

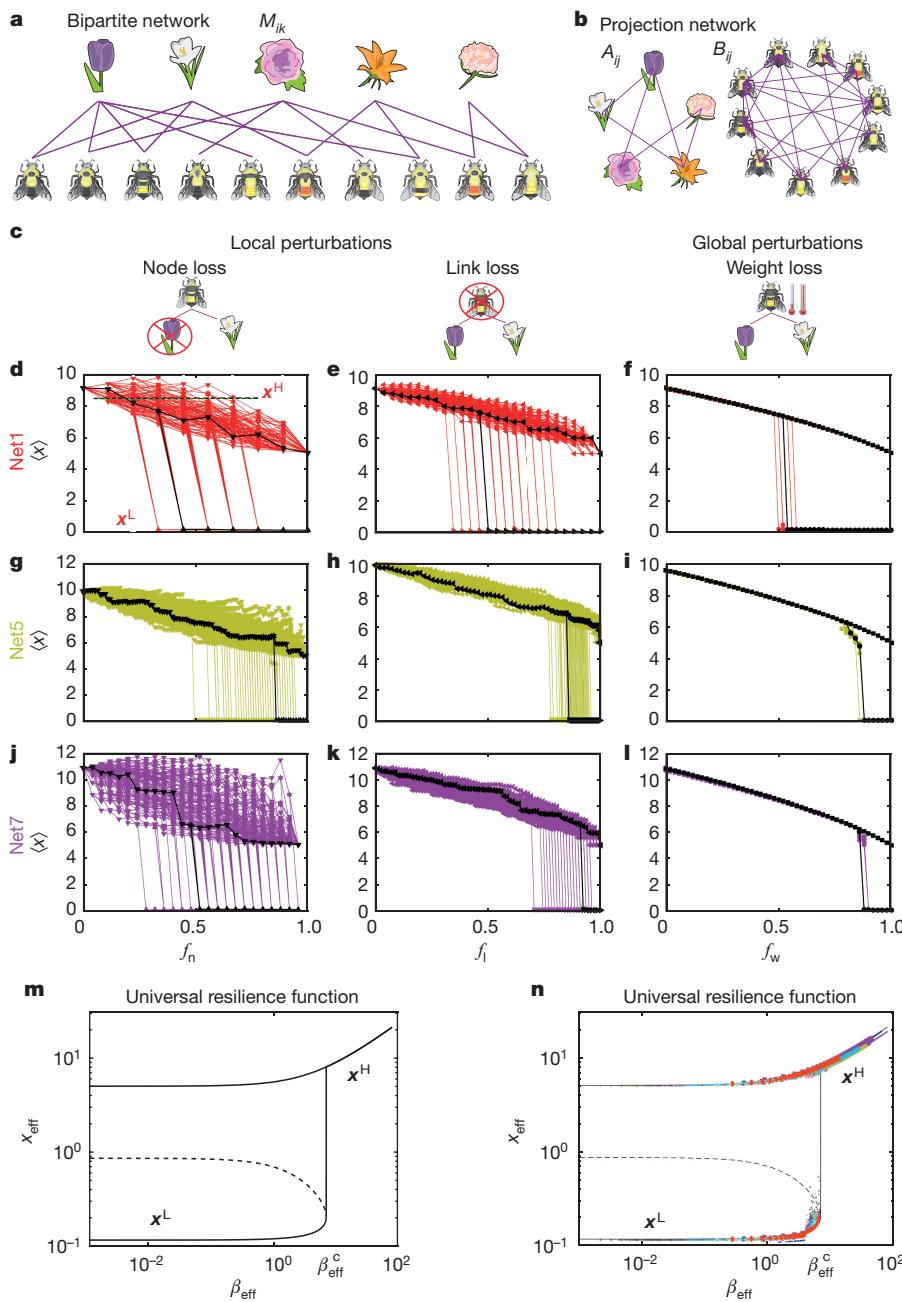


Figure 2 | Resilience in ecological networks.

a, The bipartite network M_{ik} consists of nodes representing symbiotically connected species, such as plants and pollinators or fish and anemone. **b**, From M_{ik} we construct two mutualistic networks (Supplementary Information section II.B) by linking pairs of plants that share mutual pollinators (A_{ij}), or pollinators that share mutual plants (B_{ij}). **c**, We tested the resilience of fourteen mutualistic networks (equation (5); $B_i = B = 0.1$, $C_i = C = 1$, $K_i = K = 5$, $D_i = D = 5$, $E_i = E = 0.9$ and $H_i = H = 0.1$) against: (1) extinction of a fraction f_n of plants; (2) extinction of a fraction f_i of pollinators, impacting the relevant plant link weights; (3) decreasing all weights on average to a fraction f_w of their original value, simulating a global change in the environmental conditions, for example, varying temperature. **d**, The average abundance of fish in Net1 versus f_n across 100 realizations (we highlight one of these realizations in black). At a critical fraction f_n the system undergoes a bifurcation, where in addition to the high-abundance state (x^H) an undesired low-abundance state emerges (x^L). However, owing to the multi-dimensionality of A_{ij} , each realization is microscopically distinct, and as a result the bifurcation point is unpredictable, ranging from $f_n = 0.3$ to $f_n = 0.7$ across different realizations. **e**, **f**, Similar diversity characterizes the system's response to link perturbation f_i (**e**) and global perturbations f_w (**f**). **g–l**, The difficulty in predictability is also observed in the larger networks Net5 and Net7. Additional results appear in Supplementary Figs 2–4. **m**, Our formalism predicts that in β -space the resilience function takes a universal form, similar to that shown in Fig. 1a, with bifurcation at $\beta_{\text{eff}}^c = 6.97$, independent of A_{ij} . **n**, All data in **d–l** and in Supplementary Figs 2–4, comprising 28 highly diverse networks, collapses onto the universal resilience function predicted in **m**, indicating that regardless of the network structure and the form of perturbation, the state of the system is fully determined by β_{eff} (equation (8)).

multi-dimensional equation (4) to an effective 1D equation, written in terms of x_{eff}

$$\frac{dx_{\text{eff}}}{dt} = F(x_{\text{eff}}) + \beta_{\text{eff}} G(x_{\text{eff}}, x_{\text{eff}}) \quad (7)$$

where

$$\beta_{\text{eff}} = \frac{\mathbf{I}^\top \mathbf{A} \mathbf{s}^{\text{in}}}{\mathbf{I}^\top \mathbf{A} \mathbf{I}} = \frac{\langle s^{\text{out}} s^{\text{in}} \rangle}{\langle s \rangle} \quad (8)$$

averages over the product of the outgoing and incoming degrees of all nodes. This reduction maps the multi-dimensional complex system (4) into an effective 1D equation of the form of equation (1), where

$$f(\beta_{\text{eff}}, x_{\text{eff}}) = F(x_{\text{eff}}) + \beta_{\text{eff}} G(x_{\text{eff}}, x_{\text{eff}}) \quad (9)$$

We predict, therefore, that the N^2 parameters of the microscopic description A_{ij} collapse into a single macroscopic resilience parameter β_{eff} (see equation (8)), so that regardless of the microscopic details

of the perturbation (node/link removal, weight reduction or any combination thereof), its impact on the state of the system is fully accounted for by the corresponding changes in β_{eff} . This implies that the rather diverse and unpredictable behaviours observed in Fig. 2 are, in fact, drawn from a single universal resilience function, independent of the network topology A_{ij} , and uniquely determined by the system's dynamics $F(x_i)$ and $G(x_i, x_j)$. The network A_{ij} , which is fully condensed into the single macroscopic parameter β_{eff} , determines only the specific state of the system along this function. Such mapping of equation (4) to the 1D equation (7) allows us to take advantage of the theoretical tools developed for low-dimensional systems and apply them to a broad range of complex systems.

To illustrate the power of our formalism we apply it to the mutualistic networks of Fig. 2. Reducing the multi-dimensional equation (5) to the form of equation (9) we arrive at the 1D equation (Supplementary Information section II)

$$\frac{dx_{\text{eff}}}{dt} = B + x_{\text{eff}} \left(1 - \frac{x_{\text{eff}}}{K} \right) \left(\frac{x_{\text{eff}}}{C} - 1 \right) + \beta_{\text{eff}} \frac{x_{\text{eff}}^2}{D + (E + H)x_{\text{eff}}} \quad (10)$$

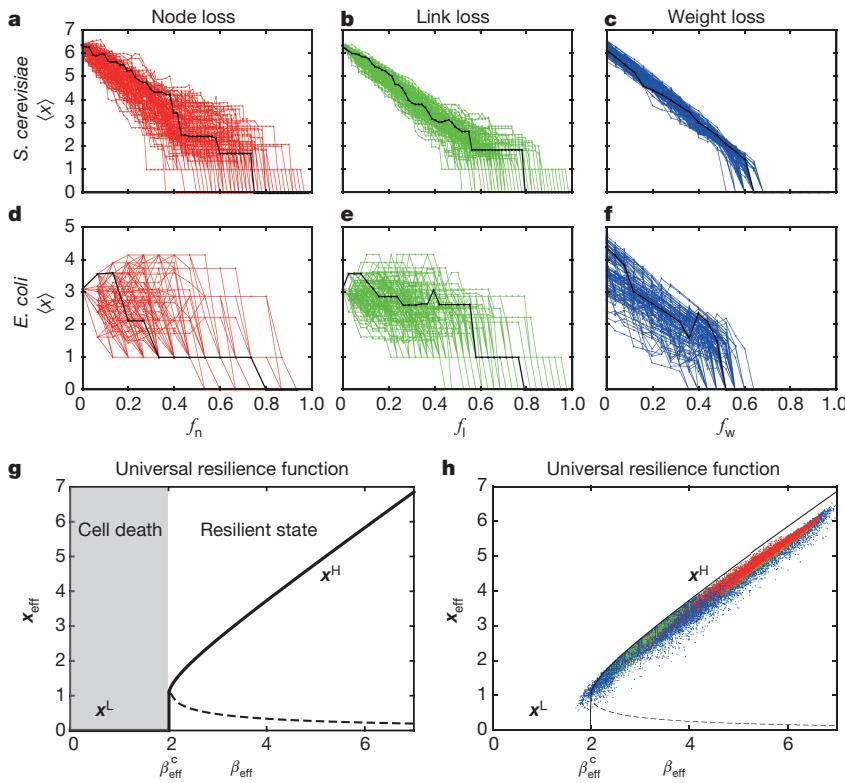


Figure 3 | Resilience in gene regulatory networks. We ran Michaelis–Menten dynamics (equation (11)) on the transcription regulatory networks of *S. cerevisiae*²⁴ and *E. coli*²⁵ (Supplementary Table 2) to model the dynamics of genetic regulation, providing the average activity $\langle x \rangle$ of all genes. **a**, By removing a sufficiently high fraction f_n of nodes the system undergoes a transition from $x^H > 0$ (alive) to $x^L = 0$ (cell death). The transition point occurs at a different value of f_n in each realization. **b, c**, Similar results are found for link perturbations and global weight changes. **d–f**, The same diversity is observed for the *E. coli* network as well. **g**, Our formalism predicts that the behaviour of gene regulation is captured by a universal resilience function of the form of Fig. 1b, with a single first-order transition at $\beta_{\text{eff}}^c = 2$. This function is determined by the regulatory dynamics of equation (11), and is independent of the network structure or the nature of its perturbations. **h**, The results of a–f shown in β -space. Regardless of the system's microscopic details, all observed data points, taken from *S. cerevisiae* or *E. coli*, induced by node perturbations (red), link perturbations (green) or weight changes (blue), collapse onto the same curve, well approximated by the analytically predicted universal resilience function (solid line).

Using equations (2) and (3), equation (10) predicts a bifurcating resilience function of the form of Fig. 1a, and a transition from a resilient state with a single stable fixed point, x^H , to a non-resilient state in which both x^H and x^L are stable. The critical point of this bifurcation is predicted to be $\beta_{\text{eff}}^c = 6.97$, a value fully determined by the dynamics, independent of the network topology A_{ij} (Fig. 2m and Supplementary Information section II).

Our formalism predicts that the diversity observed in Fig. 2d–l and in Supplementary Figs 2–4, is, in fact, driven by the universal curve of Fig. 2m. This universality can be exposed by transitioning to the natural parameter space of x_{eff} (see equation (6)) and β_{eff} (see equation (8)). Hence we re-plotted all the data of Fig. 2d–l (and Supplementary Figs 2–4, 28 diverse networks in total), in this effective β -space, finding that, as predicted, all data points collapse into a single universal curve, regardless of the size and the topology of the specific ecological network or the nature of the applied perturbation (Fig. 2n). This collapse indicates that our analytically predicted resilience function exposes a universality sustained across networks of different size, density, degree and weight distributions. Additional extensive testing of this universality appears in Supplementary Information section V.

In summary, the resilience pattern of a complex system is effectively unpredictable in the natural (x, A_{ij}) state parameter space. Once, however, we map the system into β -space we can accurately predict the system's response to diverse perturbations and correctly identify the critical points where the system loses its resilience.

Next we explore the resilience of gene regulatory networks governed by the Michaelis–Menten equation¹⁴

$$\frac{dx_i}{dt} = -Bx_i^f + \sum_{j=1}^N A_{ij} \frac{x_j^h}{x_j^h + 1} \quad (11)$$

The first term on the right-hand side of equation (11) describes degradation ($f=1$) or dimerization ($f=2$); the second term captures genetic activation, where the Hill coefficient h describes the level of cooperation in gene regulation¹⁴. Applying this model with $B=1$, $f=1$ and $h=1$ to the transcription networks of *Saccharomyces cerevisiae*²⁴ and *Escherichia coli*²⁵, we find that under sufficiently large perturbations

the cell undergoes a transition from a resilient state ($\langle x \rangle > 0$) to cell death ($\langle x \rangle = 0$, Fig. 3a–f). Once again, resilience loss strongly depends on the nature of the perturbation (gene knockout, suppression of regulatory interactions, environmental change), as well as on the differences between the wiring diagrams of *S. cerevisiae* versus *E. coli* (Supplementary Table 2). Rewriting equation (11) in the reduced form of equation (7), we find that regulatory dynamics follow the universal resilience function shown in Fig. 1b, featuring a single first-order transition from the active state to cell death at (Supplementary Information section III)

$$\beta_{\text{eff}}^c = \frac{Bh}{h-f} \left(\frac{h-f}{f} \right)^{\frac{f}{h}} \quad (12)$$

Indeed, in β -space, all trajectories of Fig. 3a–f collapse onto the analytically derived resilience function, as predicted by our formalism (Fig. 3h). In Supplementary Information section IV we develop another application area, demonstrating how to apply our formalism to predict the resilience of energy supply in the power grid.

Although the resilience function is uniquely determined by the dynamical functions $F(x_i)$ and $G(x_i, x_j)$, the actual position of the system along this curve, capturing its momentary state, is determined by the weighted network topology A_{ij} , as expressed through β_{eff} in equation (8). This prompts us to ask what aspects of the network topology determine a system's resilience. We therefore rewrite β_{eff} as

$$\beta_{\text{eff}} = \langle s \rangle + \mathcal{SH} \quad (13)$$

where $\langle s \rangle$, \mathcal{S} and \mathcal{H} represent three characteristics of A_{ij} . The dependence on the network density $\langle s \rangle$ indicates that denser networks have a larger β_{eff} (Fig. 4a). The heterogeneity in the weighted degrees s^{in} and s^{out} is captured by $\mathcal{H} = \sigma_{\text{in}}\sigma_{\text{out}}/\langle s \rangle$, where σ_{in}^2 and σ_{out}^2 are the variance of the marginal probability density functions $P(s^{\text{in}})$ and $P(s^{\text{out}})$ respectively (Fig. 4b). Finally, the symmetry between s^{in} and s^{out} is captured by $\mathcal{S} = (\langle s^{\text{in}}s^{\text{out}} \rangle - \langle s^{\text{in}} \rangle \langle s^{\text{out}} \rangle)/(\sigma_{\text{in}}\sigma_{\text{out}})$, the in–out weighted-degree correlation coefficient. This term, $-1 \leq \mathcal{S} \leq 1$, is positive when nodes

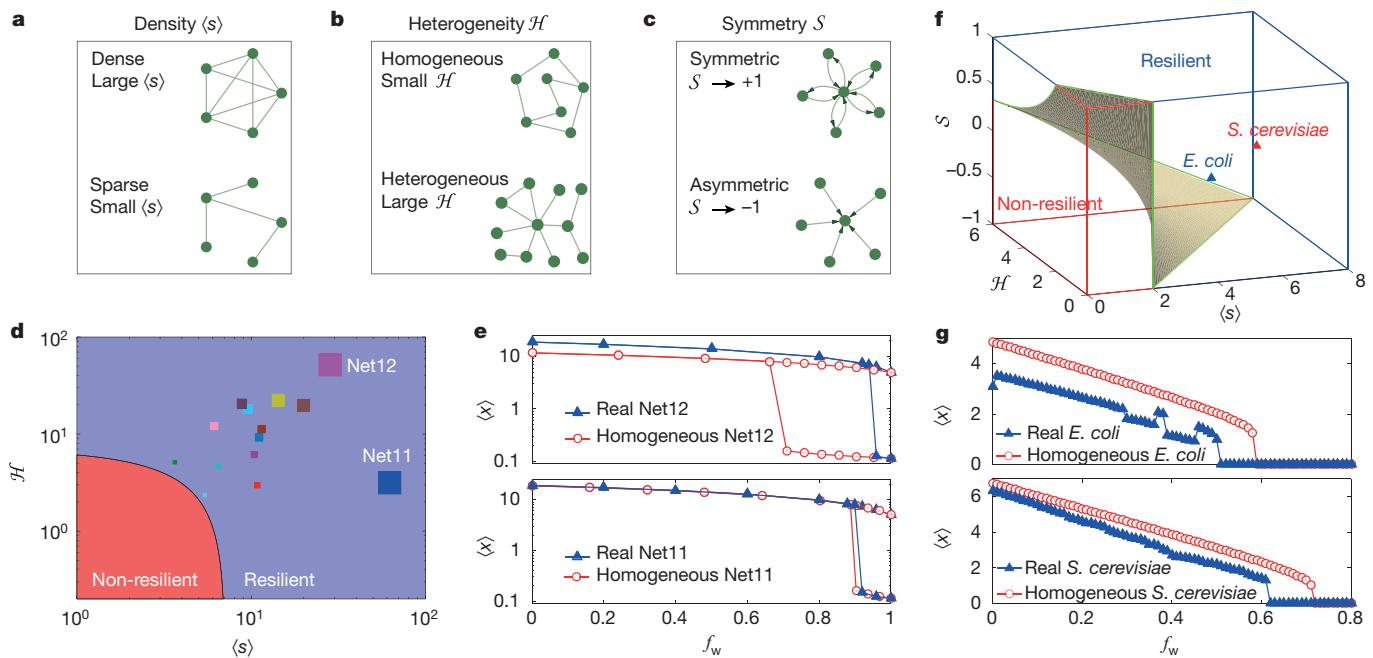


Figure 4 | The impact of A_{ij} on resilience. The topological characteristics that affect a system's resilience through β_{eff} (equation (13)) are: **a**, the network density $\langle s \rangle$; **b**, the heterogeneity in degrees or link weights \mathcal{H} ; and **c**, the symmetry \mathcal{S} , capturing the correlations between a node's in and out degrees. **d**, Phase diagram for mutualistic dynamics in the $\langle s \rangle$ – \mathcal{H} plane. In the resilient phase, the system has a single stable fixed point \mathbf{x}^{H} ; in the non-resilient phase the undesired \mathbf{x}^{L} is also stable. For this dynamics the greater β_{eff} is (square size) the deeper the system is in the resilient phase. **e**, The average state of the system $\langle x \rangle$ versus the average reduction in the link weights. For Net12, the most heterogeneous of the fourteen mutualistic networks, we observe an extreme degree of resilience, avoiding collapse up to $f_w = 97\%$ (blue triangles). A homogeneous network, with the

with large s_i^{in} tend also to have a large s_i^{out} . An undirected network $A_{ij} = A_{ji}$ is a perfectly symmetric network, as $s_i^{\text{in}} = s_i^{\text{out}}$, hence $\mathcal{S} = 1$; an asymmetric network is one where nodes with a large in-degree tend to have a small out-degree, in which case $\mathcal{S} < 0$, tending towards -1 (Fig. 4c).

Equation (13) helps us to identify the network characteristics that can enhance or weaken a system's resilience. Consider for example the resilience of the ecosystem described in equation (5). The structure of the resilience function (Fig. 2m) indicates that the greater is β_{eff} , the more resilient is the system, enduring larger perturbations before reaching the bifurcation at β_{eff}^c and risking a transition to the undesired \mathbf{x}^{L} . Since mutualistic networks are symmetric we have $\mathcal{S} = 1$ and $\mathcal{H} = \sigma_{\text{in}}^2 / \langle s \rangle \equiv \sigma^2 / \langle s \rangle$, obtaining $\beta_{\text{eff}} = \langle s \rangle + \sigma^2 / \langle s \rangle$. Hence, equation (13) increases both with density $\langle s \rangle$ and with heterogeneity \mathcal{H} giving rise to a 2D phase space: a resilient phase above the phase boundary $\langle s \rangle + \mathcal{H} = \beta_{\text{eff}}^c$, and a non-resilient phase below it. In Fig. 4d we show all fourteen mutualistic networks and their location on the $\langle s \rangle$ – \mathcal{H} phase diagram, characterizing the source of each network's resilience. For instance, Net11 and Net12 have comparable β_{eff} values and hence comparable levels of resilience, indicating that both can withstand comparable levels of perturbation before crossing the bifurcation at β_{eff}^c . However, Fig. 4d shows that while the source of Net11's resilience is its high density $\langle s \rangle$, the source of Net12's resilience is its high heterogeneity \mathcal{H} . To test this we constructed two homogeneous networks with the same densities as Net11 and Net12, but with $\mathcal{H} = 0$. As expected for Net12, reducing heterogeneity greatly decreased resilience (by $\sim 30\%$), while for Net11, whose source of resilience is density rather than heterogeneity, eliminating \mathcal{H} had only a negligible impact (Fig. 4d, e).

In regulatory dynamics the phase diagram has two domains, corresponding to an active phase and cell death. Here resilience increases

same density and $\mathcal{H} = 0$ loses resilience at $f_w = 66\%$ (red circles); hence \mathcal{H} is the source of Net12's exceptional resilience. As indicated in **d**, Net11's resilience is rooted in its high density $\langle s \rangle$. Indeed, we find that for Net11, both the original (blue triangles) and the homogeneous (red circles) networks feature comparable levels of resilience, indicating that \mathcal{H} has a marginal contribution. **f**, The phase diagram for directed transcription regulatory networks is 3D– $\langle s \rangle$, \mathcal{H} and \mathcal{S} —with the first-order transition from a living cell to cell death occurring at $\langle s \rangle + \mathcal{S}\mathcal{H} = 2$. The position of the *S. cerevisiae* and the *E. coli* networks is also shown (triangles). **g**, For both organisms $\mathcal{S} < 0$, and hence heterogeneity decreases their resilience: indeed, the homogeneous networks (red circles) withstand larger perturbations than the original networks (blue triangles).

with β_{eff} , as the larger is β_{eff} , the deeper is the system into the active state, and farther from the critical transition at β_{eff}^c (equation (12)). Since A_{ij} is directed, \mathcal{H} can both increase or decrease β_{eff} , depending on the symmetry \mathcal{S} in equation (13). Consequently, resilience is governed by three topological characteristics, where dense, symmetric and heterogeneous networks are most resilient (large β_{eff}), and sparse, antisymmetric and heterogeneous networks are least resilient (small β_{eff}). The phase diagram in the $\langle s \rangle$ – \mathcal{H} – \mathcal{S} space is shown in Fig. 4f, and the transition between states occurs along the plane $\langle s \rangle + \mathcal{S}\mathcal{H} = \beta_{\text{eff}}^c$. For the two regulatory networks we measured $\mathcal{S} = -0.083$ (for *S. cerevisiae*) and $\mathcal{S} = -0.2464$ (for *E. coli*), both negative. Hence here \mathcal{H} has a negative contribution to resilience and a homogeneous network with $\mathcal{H} = 0$ would, in fact, be more resilient. To test this prediction we compared the resilience function of the empirical networks (*E. coli* and *S. cerevisiae*) with that of the equivalent homogeneous networks in which $\langle s \rangle$ is preserved and $\mathcal{H} = 0$ (Fig. 4g). Indeed, we find that eliminating heterogeneity increases the system's resilience.

Complex systems are characterized by an inherently multi-dimensional parameter space, giving rise to diverse and unpredictable behaviour. Here, by reverting to the natural parameter space (β -space) we exposed the hidden universal patterns of network resilience. The origin of this universality is in the separation of the system's dynamics and topology. Indeed, in most systems the intrinsic behaviour of the components and the nature of the interactions between them are invariant to perturbations¹³. Perturbations only affect the structure of the network, A_{ij} , determining who interacts with whom and how strongly. Our formalism reduces A_{ij} into an effective 1D system, showing that regardless of the specific topology and weights, or the form of perturbation, the patterns of resilience depend only on the system's intrinsic dynamics. The role of the network topology is fully captured by the

1D β_{eff} , predicting that density, heterogeneity and symmetry are the three key structural factors affecting a system's resilience. They do not alter the critical points, but instead push a system far away from these critical points, helping the system to sustain large perturbations. This separation of structure and dynamics provides us with testable predictions for the system's response to different perturbations. It also suggests potential intervention strategies to avoid the loss of resilience^{26–28}, or design principles for optimal²⁹ resilient systems that can successfully cope with perturbations³⁰.

Received 13 July; accepted 14 December 2015.

- Cohen, R., Erez, K., Ben-Avraham, D. & Havlin, S. Resilience of the internet to random breakdown. *Phys. Rev. Lett.* **85**, 4626–4628 (2000).
- Venegas, J. G. et al. Self-organized patchiness in asthma as a prelude to catastrophic shifts. *Nature* **434**, 777–782 (2005).
- Perrings, C. Resilience in the dynamics of economy-environment systems. *Environ. Resour. Econ.* **11**, 503–520 (1998).
- May, R. M. Thresholds and breakpoints in ecosystems with a multiplicity of stable states. *Nature* **269**, 471–477 (1977).
- Lyapunov, A. M. The general problem of the stability of motion. *Int. J. Control.* **55**, 531–534 (1992).
- Barzel, B. & Biham, O. Quantifying the connectivity of a network: the network correlation function method. *Phys. Rev. E* **80**, 046104 (2009).
- Sole, R. V. & Montoya, M. Complexity and fragility in ecological networks. *Proc. R. Soc. Lond. B* **268**, 2039–2045 (2001).
- May, R. M., Levin, S. A. & Sugihara, G. Complex systems: Ecology for bankers. *Nature* **451**, 893–895 (2008).
- Scheffer, M. et al. Anticipating critical transitions. *Science* **338**, 344–348 (2012).
- Albert, R. & Barabási, A.-L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97 (2002).
- Ben-Naim, E., Frauenfelder, H. & Toroczkai, Z. *Complex Networks* Vol. 650 (Springer Science & Business Media, 2004).
- Barzel, B. & Barabási, A.-L. Universality in network dynamics. *Nature Phys.* **9**, 673–681 (2013).
- Barzel, B., Liu, Y.-Y. & Barabási, A.-L. Constructing minimal models for complex system dynamics. *Nature Commun.* **6**, 7186 (2015).
- Alon, U. *An Introduction to Systems Biology: Design Principles of Biological Circuits* (CRC Press, 2006).
- Berlow, E. L. et al. Simple prediction of interaction strengths in complex food webs. *Proc. Natl Acad. Sci. USA* **106**, 187–191 (2009).
- Holland, J. N., DeAngelis, D. L. & Bronstein, J. L. Population dynamics and mutualism: functional responses of benefits and costs. *Am. Nat.* **159**, 231–244 (2002).
- Pastor-Satorras, R. & Vespignani, A. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **86**, 3200 (2001).
- Dunne, J. A., Williams, R. J. & Martinez, N. D. Network structure and biodiversity loss in food webs: robustness increases with connectance. *Ecol. Lett.* **5**, 558–567 (2002).
- Wilhelm, T., Behre, J. & Schuster, S. Analysis of structural robustness of metabolic networks. *Syst. Biol.* **1**, 114–120 (2004).
- McCulloch, M., Falter, J., Trotter, J. & Montagna, P. Coral resilience to ocean acidification and global warming through pH up-regulation. *Nature Clim. Change* **2**, 623–627 (2012).
- Hui, C. Carrying capacity, population equilibrium, and environment's maximal load. *Ecol. Model.* **192**, 317–320 (2006).
- Allee, W. C. et al. *Principles of Animal Ecology* Edn 1 (WB Saunders, 1949).
- Interaction Web Database http://www.nceas.ucsb.edu/interactionweb/resources.html#plant_ant (accessed 30 September 2010).
- Balaji, S., Madan Babu, M., Iyer, L., Luscombe, N. & Aravind, L. Principles of combinatorial regulation in the transcriptional regulatory network of yeast. *J. Mol. Biol.* **360**, 213–227 (2006).
- Gama-Castro, S. et al. RegulonDB (version 6.0): gene regulation model of *Escherichia coli* k-12 beyond transcription, active (experimental) annotated promoters and textpresso navigation. *Nucleic Acids Res.* **36**, D120–D124 (2008).
- Nepusz, T. & Vicsek, T. Controlling edge dynamics in complex networks. *Nature Phys.* **8**, 568–573 (2012).
- Cornelius, S. P., Kath, W. L. & Motter, A. E. Realistic control of network dynamics. *Nature Commun.* **4**, 1942 (2013).
- Majdandzic, A. et al. Spontaneous recovery in dynamical networks. *Nature Phys.* **10**, 34–38 (2013).
- Helbing, D. & Vicsek, T. Optimal self-organization. *New J. Phys.* **1**, 13 (1999).
- Cohen, R., Erez, K., Ben-Avraham, D. & Havlin, S. Breakdown of the internet under intentional attack. *Phys. Rev. Lett.* **86**, 3682–3685 (2001).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Mohan, S. E. Flynn and A. R. Ganguly for discussions. This work was supported by an Army Research Laboratories Network Science Collaborative Technology Alliance grant (ARL NS-CTA W911NF-09-2-0053), by The John Templeton Foundation: Mathematical and Physical Sciences (grant number PFI-777), by The Defense Threat Reduction Agency (basic research grant number HDTRA1-10-1-0100) and by the European Commission (grant numbers FP7317532 (MULTIPLEX) and 641191 (CIMPLEX)).

Author Contributions All authors designed and did the research. J.G. and B.B. did the analytical calculations. J.G. analysed the empirical data and did the numerical calculations. A.-L.B. and B.B. were the lead writers of the manuscript.

Author Information All code for the reproduction of the reported results can be downloaded from <https://github.com/jianxiao/NuRsE>. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.-L.B. (alb@neu.edu).

Non-classical correlations between single photons and phonons from a mechanical oscillator

Ralf Riedinger^{1*}, Sungkun Hong^{1*}, Richard A. Norte², Joshua A. Slater¹, Juying Shang³, Alexander G. Krause², Vikas Anant³, Markus Aspelmeyer¹ & Simon Gröblacher²

Interfacing a single photon with another quantum system is a key capability in modern quantum information science. It allows quantum states of matter, such as spin states of atoms^{1,2}, atomic ensembles^{3,4} or solids⁵, to be prepared and manipulated by photon counting and, in particular, to be distributed over long distances. Such light-matter interfaces have become crucial to fundamental tests of quantum physics⁶ and realizations of quantum networks⁷. Here we report non-classical correlations between single photons and phonons—the quanta of mechanical motion—from a nanomechanical resonator. We implement a full quantum protocol involving initialization of the resonator in its quantum ground state of motion and subsequent generation and read-out of correlated photon-phonon pairs. The observed violation of a Cauchy-Schwarz inequality is clear evidence for the non-classical nature of the mechanical state generated. Our results demonstrate the availability of on-chip solid-state mechanical resonators as light-matter quantum interfaces. The performance we achieved will enable studies of macroscopic quantum phenomena⁸ as well as applications in quantum communication⁹, as quantum memories¹⁰ and as quantum transducers^{11,12}.

Over the past few years, nanomechanical devices have been discussed as possible building blocks for quantum information architectures^{9,13}. Their unique feature is that they combine an engineerable solid-state platform on the nanoscale with the possibility to coherently interact with a variety of physical quantum systems including electronic or nuclear spins, single charges, and photons^{14,15}. This feature enables mechanics-based hybrid quantum systems that interconnect different, independent physical qubits through mechanical modes.

A successful implementation of such quantum transducers requires the ability to create and control quantum states of mechanical motion. The first step—the initialization of micro- and nanomechanical systems in their quantum ground state of motion—has been realized in various mechanical systems either through direct cryogenic cooling^{16,17} or laser cooling using microwave¹⁸ and optical cavity fields¹⁹. Further progress in quantum state control has mainly been limited to the domain of electromechanical devices, in which mechanical motion couples to superconducting circuits in the form of qubits and microwave cavities¹⁵. Recent achievements include single-phonon control of a micromechanical resonator by a superconducting flux qubit¹⁶, the generation of quantum entanglement between quadratures of a microwave cavity field and micromechanical motion²⁰, and the preparation of quantum squeezed micromechanical states²¹.

Interfacing mechanics with optical photons in the quantum regime is highly desirable because it adds important features such as the ability to transfer mechanical excitations over long distances^{9,22}. In addition, the available toolbox of single-photon generation and detection allows for remote quantum state control⁷. However, micro- and nanomechanical quantum control through single optical photons has not

yet been demonstrated. One of the outstanding challenges is to achieve single-particle coupling rates that are sufficiently large to alleviate effects of optical and mechanical decoherence in the system, that is, single-photon strong co-operativity. Some of the largest optomechanical couplings have been reported in nanomechanical photonic crystal cavities²³, but are still two orders of magnitude short of that regime. Although low coupling rates can be overcome in principle by a strong and detuned coherent drive field¹⁵, such measures typically result in unwanted heating of the mechanical device (see Methods).

Here we take a different approach that allows us to circumvent these problems and to realize quantum control of single phonons through single optical photons. We use a probabilistic scheme based on the well-known DLCZ protocol (Duan, Lukin, Cirac and Zoller)²⁴, which, in its original form, uses Raman scattering for efficient generation and read-out of collective spin states of atomic ensembles. In essence, the scheme generates entanglement through single-photon interference and post-selection, which does not require strong coupling²⁵. In the context of mechanical quanta, this protocol has been used in an experiment to entangle high-frequency (40 THz) optical phonons of two bulk diamond lattices²⁶. However, the small interaction and coherence times of such phonons are incompatible with their use in quantum transduction and storage, and so it is necessary to take this approach to the level of chip-scale optomechanical systems. In addition, we minimize absorption heating by using short optical pulses in a cryogenic environment¹⁷. The combination of these techniques allows us to overcome the previous limitations and realize a photon-phonon quantum interface.

Our experiment complements previous work on single- and two-mode (opto-)mechanical squeezing in microwave circuits^{20,21}. Although these experiments were based on the same underlying interactions, they involved homodyne or heterodyne detection of light to access continuous-variable degrees of freedom of a quantum state—specifically, quadrature fluctuations in the mechanical and optical canonical variables. In contrast, the DLCZ scheme uses photon counting, which allows access to discrete quantum variables—here, in form of energy eigenstates (phonons) of the mechanical motion—and thereby enables realistic architectures for entanglement distribution and quantum networking⁷.

The mechanical system studied here is a micro-fabricated silicon photonic crystal nanobeam structure (Fig. 1a). Such optomechanical crystals co-localize optical and mechanical modes and couple them via a combination of radiation pressure and photostriction¹⁵. Our device exhibits an optical cavity resonance at wavelength $\lambda_c = 1,556$ nm and a mechanical breathing mode at frequency $\omega_m/(2\pi) = 5.3$ GHz. The cavity decay rate (full-width at half-maximum, FWHM) is $\kappa_c/(2\pi) = 1.3$ GHz and the mechanical quality factor at cryogenic temperature is $Q_m = 1.1 \times 10^6$ (see Methods). Pulsed optical driving at laser frequency $\omega_L = \omega_c \pm \omega_m$ (in which $\omega_c = 2\pi c/\lambda_c$ is the cavity frequency and c is the vacuum speed of light) allows to realize two different types of

¹Vienna Center for Quantum Science and Technology (VCQ), Faculty of Physics, University of Vienna, A-1090 Vienna, Austria. ²Kavli Institute of Nanoscience, Delft University of Technology, 2628CJ Delft, The Netherlands. ³Photon Spot Inc., Monrovia, California 91016, USA.

*These authors contributed equally to this work.

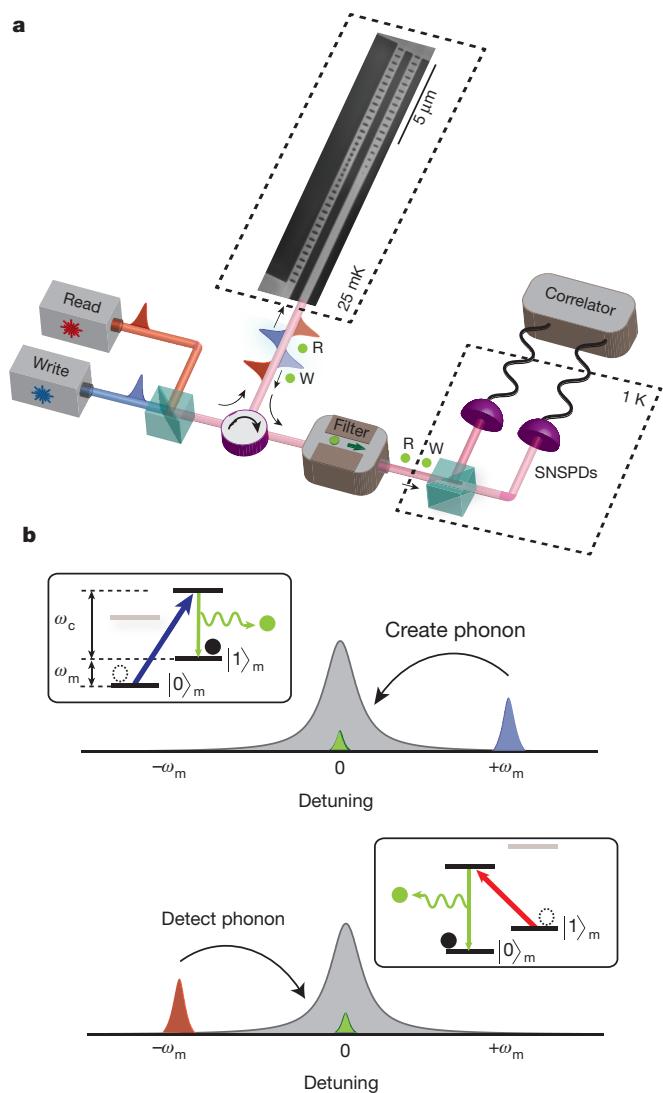


Figure 1 | Generation and read-out of photon-phonon pairs. **a**, Schematic of the experiment. Two independent lasers (stabilized to a wave-meter) are used to generate a sequence of 'write' and 'read' pulses with tunable time delay δt . They are sent through a circulator and drive a nanomechanical photonic crystal cavity (a scanning electron microscope image of which is shown in the inset) that is mounted inside a dilution refrigerator at a base temperature of 25 mK, which prepares the device in its quantum ground state of motion. For each pulse, Stokes and anti-Stokes Raman scattering creates single photons (green dots) from the write (W) and the read (R) pulse, respectively, that are emitted at a frequency ω_c . The detuned pump fields are strongly suppressed by optical filtering and only the Raman scattered photons are measured by two superconducting nanowire single-photon detectors (SNSPDs) in the output ports of a 50/50 beam-splitter. The time of each photon detection event is recorded and is then correlated in post-processing to obtain both auto- and cross-correlations of the emitted photons. A more detailed explanation of the experimental set-up is provided in Methods. **b**, Pulsed optomechanical interactions in frequency space. A blue-detuned write pulse realizes a two-mode squeezing interaction (blue and green pulses; see text). Cavity-enhanced Stokes Raman scattering generates a single phonon, stored as excitation on the mechanical resonator, and a single (W) photon, which is emitted from the cavity on resonance (upper panel). Reading out of the phonon utilizes a red-detuned read pulse, which swaps the mechanical excitation onto the optical cavity field, hence creating a single (R) photon (lower panel). The insets depict the relevant energy level diagrams for the two processes, reminiscent of the Λ -schemes in atomic Raman scattering. The grey bars indicate the energy levels that are not involved in the depicted process (Stokes or anti-Stokes), but in the other one.

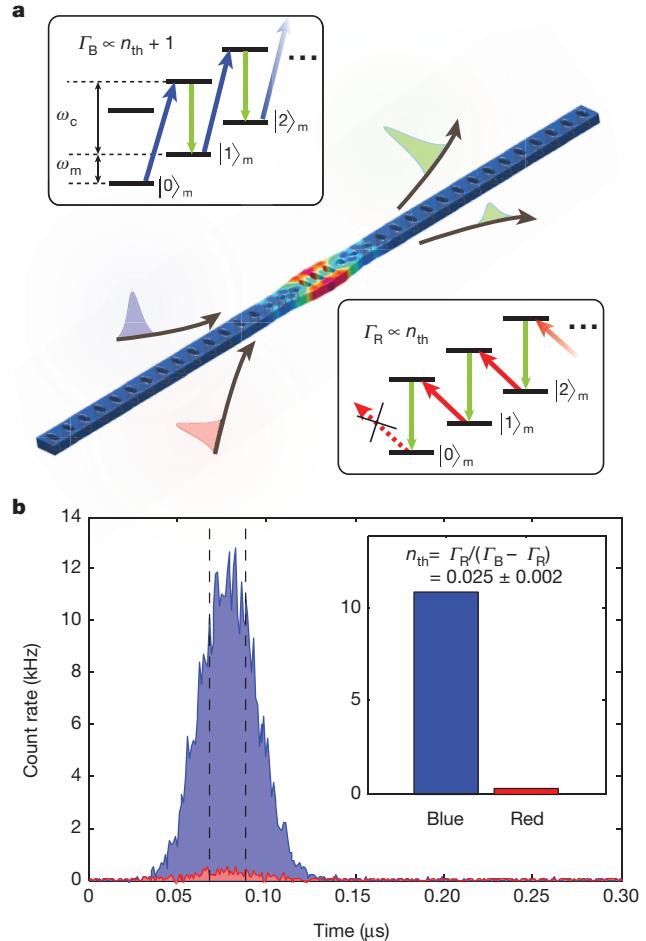


Figure 2 | Mechanical quantum ground state preparation. **a**, Principle of sideband thermometry. The finite element method simulation depicted in the main panel shows the structure of the mechanical breathing mode under investigation. The upper (lower) inset shows the energy level scheme in case of blue- (red-) detuned pumping and the resultant cavity-enhanced Stokes (anti-Stokes) scattering. The corresponding scattering rates Γ_R and Γ_B are proportional to thermal occupation of the mechanics n_{th} and $n_{th} + 1$, respectively, and hence show a strong asymmetry when the mechanics are close to the quantum ground state. **b**, Sideband asymmetry. The optomechanical device is pumped with a sequence of alternating blue- and red-detuned optical pulses at frequency $\omega_c \pm \omega_m$ (optical energy per pulse $E_{opt} = 33$ fJ; FWHM of 28.4 ns; 500- μ s separation of pulse sequences). Shown are the count rates recorded by the SNSPDs as a function of the arrival time of the scattered photons (blue, blue-detuned pulse; red, red-detuned pulse). This data has been corrected for leakage of pump photons through the optical filters, which was independently measured and subtracted from our data (see Methods). The inset shows a histogram of the total counts that are obtained when averaging over a 20-ns window centred on the peak (within the dashed lines). The pronounced asymmetry in the rates (of more than a factor of 40) corresponds to a thermal occupancy of $n_{th} = \Gamma_R / (\Gamma_B - \Gamma_R) = 0.025 \pm 0.002$ and to a mode temperature of 69 mK.

interactions on the basis of cavity-enhanced Stokes (+) and anti-Stokes (−) Raman scattering (Fig. 1b). A blue-detuned pulse ($\omega_L = \omega_c + \omega_m$) results in two-mode squeezing with interaction Hamiltonian $H_{tms} \propto \hbar g_0 (\hat{a}_m^\dagger \hat{a}_o^\dagger + \hat{a}_m \hat{a}_o)$, in which \hat{a}_m^\dagger and \hat{a}_o^\dagger are the creation (annihilation) operators of the mechanical and optical mode, respectively, g_0 is the effective optomechanical coupling rate (here, $g_0/(2\pi) = 825$ kHz; see Methods) and \hbar is the reduced Planck constant. This interaction generates photon-phonon pairs in close analogy to the photon-photon pairs generated in parametric down-conversion²⁷. A red-detuned pulse ($\omega_L = \omega_c - \omega_m$) allows read-out of the mechanical state through the optomechanical beam-splitter interaction

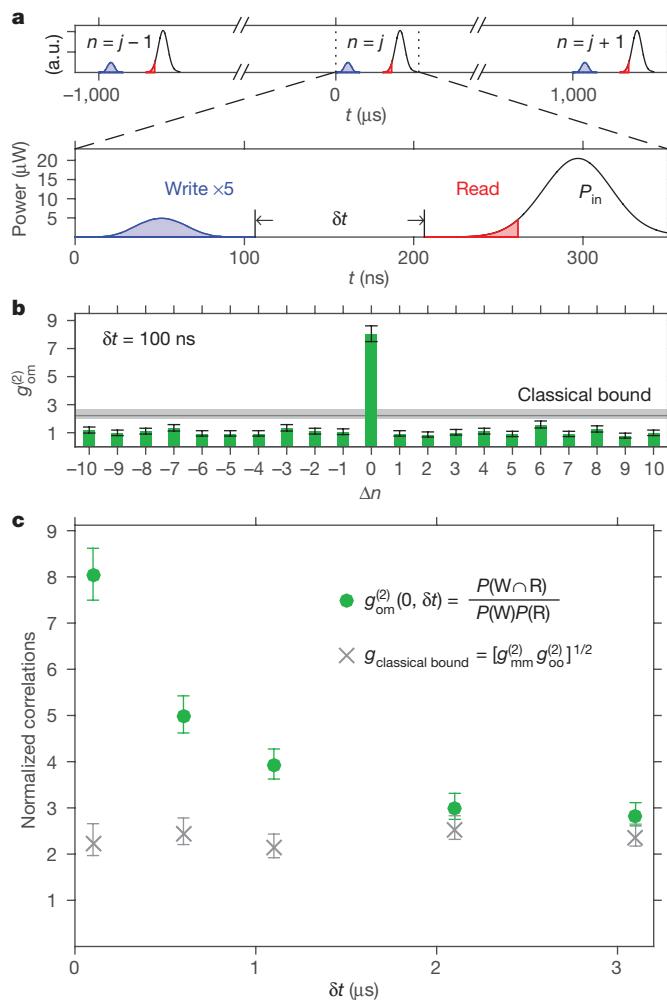


Figure 3 | Non-classical photon-phonon correlations. **a**, Driving pulse sequence. A pair of one write (blue) and one read (red) pulse is sent to the device every 1 ms. The long idle phase between pulse pairs ensures the ground-state initialization by cryogenic cooling. Each pulse sequence is labelled with a number (n). The read pulse is delayed by δt with respect to the write pulse, and only the first 55 ns, equivalent to a read-pulse power of about 50 fJ, are used for the data evaluation. This reduces the influence of absorption heating while maintaining reasonable state swap fidelity. **b**, Violating a Cauchy–Schwarz inequality. Shown is the cross-correlation (green bars) between the mechanical (read pulse) and optical state (write pulse) for $\delta t = 100$ ns, as well as the classical (Cauchy–Schwarz) bound obtained from the autocorrelations at $\Delta n = 0$ (grey horizontal line, shading indicates a 68% confidence interval; see text). For photon–phonon pairs that emerge from different pulse sequences ($\Delta n \neq 0$) the Cauchy–Schwarz inequality is fulfilled, $\langle g_{\text{om}}^{(2)}(\Delta n \neq 0, 100 \text{ ns}) \rangle = 1.04 \pm 0.04$, consistent with statistical independence. For pulses from the same pair, the cross-correlation $g_{\text{om}}^{(2)}(0, 100 \text{ ns})$ clearly exceeds the classical bound. $g_{\text{om}}^{(2)}$ can be interpreted as the ratio of heralded phonons n_h to unheralded (thermal) phonons n_{th} at the time of the read pulse. **c**, Storage of non-classical correlations. Shown is the dependence of the cross-correlation on the time delay δt between the write and read pulses. For each data point, the classical bound is measured independently through the normalized autocorrelation functions of the write (W) and read (R) photons. For increasing δt , the photon–phonon cross-correlations decrease, but stay above the classical limit even beyond 1 μs. The main contribution to the loss of correlation is heating by absorption of the write pulse (see Methods). All error bars represent a 68% confidence interval.

$H_{\text{bs}} \propto \hbar g_0 (\hat{a}_m \hat{a}_o^\dagger + \hat{a}_m^\dagger \hat{a}_o)$, in which an anti-Stokes scattering event realizes a state swap between the mechanical and optical cavity mode.

Our protocol consists of three distinctive steps. First, we initialize the mechanical system in its quantum ground state of motion by cryogenic cooling. Second, a short blue pulse creates a photon–phonon pair

and leaves the originally empty mechanical and optical modes $|0_m\rangle$ and $|0_o\rangle$ at frequencies ω_m and ω_c , respectively, in the state $|\Phi_{\text{om}}\rangle = |00\rangle + \sqrt{p}|11\rangle + p|22\rangle + \mathcal{O}(p^{3/2})$. Here p is the probability for a single Stokes scattering event to take place. Residual heating through optical absorption introduces additional noise to the state (see Methods). Finally, a strong red pulse is used to read out the phonon state via emission of an anti-Stokes scattered photon²⁸. We confirm the non-classical photon–phonon correlations on the basis of an observed violation of a Cauchy–Schwarz inequality for the cross-correlation of the coincidence measurements between the Stokes and anti-Stokes photons³.

Precooling of the nanomechanical device is performed using a dilution refrigerator that operates at a base temperature of approximately 25 mK. If the mechanical system is in its quantum ground state of motion, then anti-Stokes processes cannot occur because no additional phonons can be extracted to support the scattering. This is in contrast to Stokes processes, which deposit mechanical energy and hence can always occur. As a consequence, the asymmetry in the scattering rates of these two processes is a direct measurement of the mean thermal phonon occupancy n_{th} . Using such photon-counting based sideband thermometry¹⁷, we find $n_{\text{th}} \lesssim 0.025$ (see Fig. 2).

We create the desired photon–phonon pairs using a blue-detuned ‘write’ pulse that is sufficiently weak to minimize the effects of residual absorption heating (FWHM, 28.4 ns; energy, 40 fJ). We find the relevant probability to generate a Stokes scattered photon on cavity resonance to be $p \approx 3.0\%$. Subsequently, a red-detuned ‘read’ pulse (effective length, 55 ns; energy of approximately 50 fJ) is injected at a time delay δt (see Fig. 3a), resulting in a phonon-to-photon conversion efficiency of approximately 3.7% (see Methods).

We correlate the measured Stokes- and anti-Stokes photons via the cross-correlation function $g_{\text{om}}^{(2)}(\Delta n, \delta t) = P(W \cap R) / [P(W)P(R)]$, which is computed for read and write pulses originating from pulse sequences from different trials separated by Δn iterations (see Fig. 3). $P(W \cap R)$ is the probability for a joint detection of both a Stokes (W, ‘write’) and an anti-Stokes (R, ‘read’) photon from these pulses, and $P(W)$ and $P(R)$ are the unconditional probabilities to detect either of the two photons. For all pair correlations of classical origin, the value of $g_{\text{om}}^{(2)}$ is bounded by a Cauchy–Schwarz inequality of the form³ $g_{\text{om}}^{(2)}(0, \delta t) \leq [g_{\text{oo}, \delta t}^{(2)}(0)g_{\text{mm}, \delta t}^{(2)}(0)]^{1/2}$, in which $g_{\text{oo}, \delta t}^{(2)}(0)$ and $g_{\text{mm}, \delta t}^{(2)}(0)$ are the autocorrelation functions for the optical and mechanical mode, respectively (see Methods). A violation of this inequality^{3,29,30} is an unambiguous measure for the non-classicality of the generated photon–phonon state. The Cauchy–Schwarz inequality for coincidence detection marks a well-defined border between the quantum and classical domain. It is based on the fact that the Glauber–Sudarshan phase-space function, or P function, is positive definite for every classical field. This places a fundamental limit on the relative strength of measurable cross-correlations versus autocorrelations between classical fields. Previous applications of this limit include the distinction between the classical and quantum field theoretical predictions for the photoelectric effect²⁹, and the storage and retrieval of non-classical states in the collective emission from an atomic ensemble³. A detailed derivation of the Cauchy–Schwarz inequality for the case of non-stationary fields, as are being used here, is provided in ref. 3.

We find a clear violation for an extended regime of time delays. Figure 3b shows the value of g_{om} at a time delay of 100 ns. For pairs emitted from the same pulse sequence ($\Delta n = 0$) we find that

$$g_{\text{om}}^{(2)}(0, 100 \text{ ns}) = 8.0 + 0.6 - 0.5$$

and

$$[g_{\text{oo}, 100 \text{ ns}}^{(2)}(0)g_{\text{mm}, 100 \text{ ns}}^{(2)}(0)]^{1/2} = 2.09 + 0.23 - 0.16$$

so

$$g_{\text{om}}^{(2)}(0, 100 \text{ ns}) \not\leq [g_{\text{oo}, 100 \text{ ns}}^{(2)}(0)g_{\text{mm}, 100 \text{ ns}}^{(2)}(0)]^{1/2}$$

which obviously violates the classical bound. As expected, pairs emitted from different pulse sequences ($\Delta n \neq 0$) are uncorrelated and hence fulfil the inequality. Upon increasing the time delay further, we find a violation even beyond $\delta t = 1 \mu\text{s}$ (see Fig. 3c), which demonstrates that we can store and retrieve non-classical states for an extended time interval. Nevertheless, the lifetime of these non-classical correlations is still much shorter than the lifetime of the mechanical excitations, $Q/\omega_m \approx 34 \mu\text{s}$. We attribute this to the fact that the dynamics are dominated by heating caused by absorption of pump photons, which after some onset time drives the mechanical system towards a thermal state (see Methods). As a consequence, reducing the energy of the write pulse further should allow non-classical correlations to be maintained for much longer times. In addition, upon further reduction of the absorption heating of the read pulse, even higher values for the cross-correlation are obtained.

The cross-correlation is also linked to the autocorrelation of the heralded mechanical state. If one considers two-mode optomechanical squeezing acting on an initial mechanical thermal state, and if $g_{\text{om}}^{(2)} \gg 1$ —as is the case in our experiment—then one obtains $g_{\text{mm,heralded}}^{(2)} \approx 4/(g_{\text{om}}^{(2)} - 1)$. The largest value for $g_{\text{om}}^{(2)}$ observed in our experiment was $g_{\text{om}}^{(2)}(0, 100 \text{ ns}) = 19.6 - 2.8 + 3.9$ (using an energy of 1.7 fJ in the first 30 ns of the read pulse; see Methods). In other words, our system should allow for a Hanbury Brown and Twiss experiment with phonons yielding $g_{\text{mm,heralded}}^{(2)} \approx 0.22$. A direct measurement of this value with the current experimental parameters is difficult without a prohibitively large number of pulse sequences.

In summary, we have demonstrated non-classical correlations between single photons and phonons from a nanomechanical resonator. This is a crucial step towards on-chip photon–phonon quantum interfaces, which are relevant for future solid-state based quantum information and communication architectures. For example, the observed photon–phonon correlation of $g_{\text{om}}^{(2)} = 19.6$ suggests that conditional mechanical Fock-state preparation should be possible with fidelities exceeding 85% (see Methods). The ability to store and retrieve non-classical states over extended storage times that we reported also shows that nano-optomechanical resonators are a promising candidate for quantum memories. The performance of the system we have demonstrated constitutes an improvement of almost two orders of magnitude on previous lifetimes of stored non-classical single-phonon states¹⁶. Finally, photon–phonon conversion on the single particle level is required to extend the ongoing efforts on mechanically transduced conversion between microwave and optical fields¹² into the quantum domain¹¹.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 November; accepted 10 December 2015.

Published online 18 January 2016.

1. Wilk, T., Webster, S. C., Kuhn, A. & Rempe, G. Single-atom single-photon quantum interface. *Science* **317**, 488–490 (2007).
2. Stute, A. *et al.* Quantum-state transfer from an ion to a photon. *Nature Photon.* **7**, 219–222 (2013).
3. Kuzmich, A. *et al.* Generation of nonclassical photon pairs for scalable quantum communication with atomic ensembles. *Nature* **423**, 731–734 (2003).
4. van der Wal, C. H. *et al.* Atomic memory for correlated photon states. *Science* **301**, 196–200 (2003).
5. Yilmaz, S. T., Fallahi, P. & Imamoglu, A. Quantum-dot-spin single-photon interface. *Phys. Rev. Lett.* **105**, 033601 (2010).

6. Hensen, B. *et al.* Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres. *Nature* **526**, 682–686 (2015).
7. Kimble, H. J. The quantum internet. *Nature* **453**, 1023–1030 (2008).
8. Romero-Isart, O. Quantum superposition of massive objects and collapse models. *Phys. Rev. A* **84**, 052121 (2011).
9. Stannigel, K., Rabl, P., Sørensen, A. S., Zoller, P. & Lukin, M. D. Optomechanical transducers for long-distance quantum communication. *Phys. Rev. Lett.* **105**, 220501 (2010).
10. Chang, D. E., Safavi-Naeini, A. H., Hafezi, M. & Painter, O. Slowing and stopping light using an optomechanical crystal array. *New J. Phys.* **13**, 023003 (2011).
11. Barzanjeh, S., Abdi, M., Milburn, G. J., Tombesi, P. & Vitali, D. Reversible optical-to-microwave quantum interface. *Phys. Rev. Lett.* **109**, 130503 (2012).
12. Bochmann, J., Vainsencher, A., Awschalom, D. D. & Cleland, A. N. Nanomechanical coupling between microwave and optical photons. *Nature Phys.* **9**, 712–716 (2013).
13. Wallquist, M., Hammerer, K., Rabl, P., Lukin, M. & Zoller, P. Hybrid quantum devices and quantum engineering. *Phys. Scr.* **T137**, 014001 (2009).
14. Poot, M. & van der Zant, H. S. J. Mechanical systems in the quantum regime. *Phys. Rep.* **511**, 273–335 (2012).
15. Aspelmeyer, M., Kippenberg, T. J. & Marquardt, F. Cavity optomechanics. *Rev. Mod. Phys.* **86**, 1391–1452 (2014).
16. O’Connell, A. D. *et al.* Quantum ground state and single-phonon control of a mechanical resonator. *Nature* **464**, 697–703 (2010).
17. Meenehan, S. M. *et al.* Pulsed excitation dynamics of an optomechanical crystal resonator near its quantum ground state of motion. *Phys. Rev. X* **5**, 041002 (2015).
18. Teufel, J. D. *et al.* Sideband cooling of micromechanical motion to the quantum ground state. *Nature* **475**, 359–363 (2011).
19. Chan, J. *et al.* Laser cooling of a nanomechanical oscillator into its quantum ground state. *Nature* **478**, 89–92 (2011).
20. Palomaki, T. A., Teufel, J. D., Simmonds, R. W. & Lehnert, K. W. Entangling mechanical motion with microwave fields. *Science* **342**, 710–713 (2013).
21. Wollman, E. E. *et al.* Quantum squeezing of motion in a mechanical resonator. *Science* **349**, 952–955 (2015).
22. Safavi-Naeini, A. H. & Painter, O. Proposal for an optomechanical traveling wave phonon–photon translator. *New J. Phys.* **13**, 013017 (2011).
23. Safavi-Naeini, A. H., Alegre, T. P. M., Winger, M. & Painter, O. Optomechanics in an ultrahigh-Q two-dimensional photonic crystal cavity. *Appl. Phys. Lett.* **97**, 181106 (2010).
24. Duan, L. M., Lukin, M. D., Cirac, J. I. & Zoller, P. Long-distance quantum communication with atomic ensembles and linear optics. *Nature* **414**, 413–418 (2001).
25. Cirillo, C., Cirac, J. I., García-Fernández, P. & Zoller, P. Creation of entangled states of distant atoms by interference. *Phys. Rev. A* **59**, 1025–1033 (1998).
26. Lee, K. C. *et al.* Entangling macroscopic diamonds at room temperature. *Science* **334**, 1253–1256 (2011).
27. Wu, L.-A., Kimble, H. J., Hall, J. L. & Wu, H. Generation of squeezed states by parametric down conversion. *Phys. Rev. Lett.* **57**, 2520–2523 (1986).
28. Cohen, J. D. *et al.* Phonon counting and intensity interferometry of a nanomechanical resonator. *Nature* **520**, 522–525 (2015).
29. Clauser, J. F. Experimental distinction between the quantum and classical field-theoretic predictions for the photoelectric effect. *Phys. Rev. D* **9**, 853–860 (1974).
30. Förtsch, M. *et al.* A versatile source of single photons for quantum information processing. *Nature Commun.* **4**, 1818 (2013).

Acknowledgements We thank K. Hammerer and S. Hofer for discussions, and T. Graziosi, J. Hill, J. Hoelscher-Obermaier, Y. Liu, L. Procopio, A. Safavi-Naeini, E. Schafler, G. Steele and W. Wieczorek for experimental support. We acknowledge assistance from the Kavli Nanolab Delft, in particular from M. Zuidam and F. Dirne. This project was supported by the European Commission (cQOM, SIQS, IQUOEMS), a Foundation for Fundamental Research on Matter (FOM) Projectruimte grant (15PR3210), the Vienna Science and Technology Fund WWTF (ICT12-049), the European Research Council (ERC CoG QLev4G), and the Austrian Science Fund (FWF) under projects F40 (SFB FOQUS) and P28172. R.R. is supported by the FWF under project W1210 (CoQuS) and is a recipient of a DOC fellowship of the Austrian Academy of Sciences at the University of Vienna.

Author Contributions All authors contributed substantially to the content of this paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.G. (s.groeblicher@tudelft.nl) and M.A. (markus.aspelmeyer@univie.ac.at).

METHODS

Device fabrication and characterization. The optomechanical device used for this experiment (see Extended Data Fig. 1) is fabricated from a silicon-on-insulator wafer, with a device layer thickness of 250 nm and 3 μ m of buried oxide. The structures are patterned using an electron beam writer and are then transferred into the top silicon layer in an SF₆/O₂ atmosphere using a reactive ion-etcher. The devices are finally released and undercut using concentrated hydrofluoric acid. We design the nanobeams such that the fundamental mechanical breathing mode is at 5.3 GHz (see Fig. 2a) and the optical resonance is around 1,550 nm (the measured wavelength for the device used here is 1,556 nm)³¹. The optical and the mechanical modes are co-localized in the centre of the beam, where we create a defect region of the photonic and phononic bandgap, allowing for an optomechanical coupling rate $g_0/(2\pi) = 825$ kHz. To minimize the thermalization time to the surrounding bath we opted, unlike previous designs, to not use any additional phononic shielding. As a consequence, the mechanical quality factors at base temperature are found to be about 1.1×10^6 (see Methods section ‘Mechanical response to optical pulses’), compared to values greater than 10^7 with a phononic shield¹⁷. The laser pulses are coupled directly into a tapered waveguide through an optical fibre with a lensed tip³², achieving efficiencies of about 60%. The optical mode of the nanobeam is evanescently coupled to the waveguide, which is terminated with a periodic array of holes, which acts as a mirror, allowing us to collect the light in reflection. For this experiment we chose a critically coupled device (internal losses equal to external losses) with an optical linewidth $\kappa_c/(2\pi)$ of approximately 1.3 GHz. This places us well within the so-called resolved-sideband regime ($\omega_m > \kappa_c$).

Set-up. In this section, we provide a detailed description of the experimental set-up. It consists of a ‘pump part’, a ‘detection part’ and the ‘electronic control part’ (see Extended Data Fig. 2).

Pump part. We use two identical, tunable continuous-wave lasers (New Focus 6728) as our light sources. The lasers are detuned and stabilized to the blue and red side, respectively, of the cavity resonance of the device (1,556.21 nm). The detuning is set to be the mechanical frequency (5.307 GHz). The two lasers separately pass through voltage-controlled tunable optical filters (MicronOptics FFP-TF2, free spectral range of about 18 GHz, bandwidth of about 50 MHz) to suppress any potential background emissions dispersed in frequency space. To create short optical pulses, we modulate the filtered continuous-wave fields using acousto-optic modulators (AOMs; IntraAction) and an additional electro-optic amplitude modulator (EOM; EO Space). We use variable optical attenuators (VOAs; Sercalo) on each path to control the pulse power. The pulses are combined on a variable optical coupler and then sent to the device in the dilution refrigerator (Vericold E21) via an optical circulator. At the device (OMC, optomechanical crystal), the optomechanical interaction with the blue (red) detuned pulses generates down-(up-) converted photons, whose frequency is in resonance with the optical cavity frequency of the device. The scattered photons are reflected back from the OMC into the optical fibre and routed to the detection part through the output port of the circulator.

Detection part. Two voltage-controlled optical filters (MicronOptics FFP-TF2, specification as above) are installed in series at the beginning of the detection path. These filters are tuned on resonance with the OMC cavity frequency such that they only allow (anti-) Stokes scattered photons to be transmitted, while strong off-resonant pump photons are rejected (suppression of about 84 dB). After the filters, a 50:50 beam splitter divides the path. Each output is additionally filtered by broadband wavelength-division multiplexors (WDM), and fibre-coupled to two superconducting nanowire single-photon detectors (SNSPD; PhotonSpot, detection efficiency of about 90%, dark count rate of <10 Hz). The SNSPDs are mounted on the 1-K plate inside the dilution refrigerator. Upon receiving a photon, the SNSPD generates a brief voltage spike, which is then electrically registered by a time-correlated single-photon counting module (TCSPC; PicoQuant TimeHarp 260 NANO). The overall efficiency of detecting a photon leaving the OMC is approximately 2.7% (see below).

Control part. To generate programmable optical pulses and to detect photons synchronously, we use a digital pulse generator (DPG; Highland Technology P400) and an arbitrary waveform generator (AWG; Agilent Technologies 81180A). We first program the DPG to generate a TTL (transistor-transistor logic) gate voltage signal for the AOM on the read (red) path and to trigger the TCSPC synchronously. The DPG additionally triggers the AWG, which then generates a TTL gate voltage for the AOM and a voltage pulse for the EOM on the write (blue) path.

Mechanical response to optical pulses. Over the past few years, several experiments have demonstrated precise control over optical and mechanical states through continuous optomechanical driving, including coherent state transfer^{20,33,34} and microwave-to-optics conversion^{12,35,36}. Owing to the unavailability of the regime of single-photon strong co-operativity, strong drive fields have to be used to achieve the desired coupling strength³⁷. This leads to unwanted

heating effects, in particular in the optical domain. Because the mechanism of optical absorption couples only indirectly to the mechanical mode of interest³², using short optical pulses as non-stationary drive fields can substantially suppress the heating on short timescales—in particular at low temperatures¹⁷.

Here, we probe the thermal response of the mechanical mode by pump-probe-type measurements; we first send a short blue-detuned pump pulse onto the OMC cavity to intentionally heat the mode, and subsequently inject a red-detuned probe pulse to read out the phonon occupancy of the mode. By repeating the experiment with varying time delay between the pump and the probe pulses, we monitor time-dependent evolution of the phonon occupancy of the mode with a fixed initial impulse heating. The time delay δt is defined as the delay between the end of the pump (blue) pulse and the start of the probe (red) pulse detection window, as indicated in Fig. 3a. In that way, the probing is performed after the optical absorption of the pump photons is completed. To ensure that the mechanical mode fully re-thermalizes to the bath, we set the duty cycle of sending another blue pulse after the red pulse to be one millisecond. To improve the signal-to-noise ratio, the pulse energies of the blue (200 fJ) and red pulses (2 pJ) used here are substantially larger than in the cross-correlation measurements.

The effective mode temperature is inferred from the average count rate observed after sending the red pulse (C_R). C_R can be decomposed into three terms: (1) the rate proportional to the (on-resonance) anti-Stokes Raman scattered pump photons (C_{AS}); (2) a term corresponding to pump photons leaked through the optical filters (C_{Leak}); and (3) an additional anti-Stokes scattering term due to heating (ref. 17) of the mode during the read-out pulse (C_{Heat}). We minimize C_{Heat} by taking into account only the first 30 ns of the red pulse as the ‘logical’ red pulse. C_{Leak} gives a constant offset to the signal. C_{AS} directly reflects the effective temperature of the mode, because the anti-Stokes scattering rate is proportional to the average number of phonons (n_m) in the mode (see Fig. 2a). To that end, we deduce that $C_R(\delta t) = C_{AS}(\delta t) + C_{Leak} = \alpha n_m(\delta t) + C_{Leak}$, in which α is the constant of proportionality.

The long-term response of the mechanical mode to the initial blue pump pulse is shown in Extended Data Fig. 3a. It exhibits an exponential decay with a time constant of $T_d = 34.4 \mu$ s, which is interpreted as the mechanical damping time. The corresponding mechanical quality factor is then $Q = \omega_m T_d \approx 1.1 \times 10^6$.

In addition, we probe the short-term response of the mechanics within one microsecond after the blue pulse in more detail (Extended Data Fig. 3b). We observe an increase in C_R with a time constant of 0.37 μ s (fit to a simple exponential curve). These data reveal slow turn-on dynamics of pulse-induced heating, as previously studied¹⁷. This time constant is even shorter than the decay of the cross-correlations (see Fig. 3c), which we attribute to the increased thermal conductivity of silicon at higher temperatures, caused by absorption of increased optical pump energies.

Characterization of the detection scheme. *Detection efficiency.* We first calibrate the fibre-to-chip coupling efficiency (η_{fc}) by sending in light far off-resonant from the OMC cavity, and then measure the reflected power ($\eta_{fc} = 60.3\%$ one-way). The device impedance ratio (η_c), that is, the ratio of external coupling losses κ_{ext} to total losses κ_c , is measured through the depth and the linewidth of the optical resonance, and is found to be $\eta_c = \kappa_{ext}/\kappa_c = 0.5$. The detection efficiency of scattered photons for each detector (η_i , $i = 1, 2$) consists of η_{fc} , η_c , the total losses of the remaining detection paths ($\eta_{path,i}$) and the quantum efficiencies of the SNSPDs ($\eta_{QE,i}$). To measure η_i pulses with calibrated energy are sent off-resonantly to the OMC (P_{in}), and the reflected photons transmitted through the optical filters are detected by the SNSPDs (P_{out}). P_{in}/P_{out} corresponds to $\eta_{fc} \times \eta_c \times \eta_{path,i} \times \eta_{QE,i}$, which we measure to be 0.013 for SNSPD1 and 0.019 for SNSPD2. Therefore, we deduce $\eta_i = \eta_c \eta_{fc} \eta_{path,i} \eta_{QE,i}$ to be $\eta_1 = 1.1\%$ and $\eta_2 = 1.6\%$. The detection efficiency of SNSPD1 (characterized quantum efficiency $\eta_{QE,1} = 65\%$) is lower than SNSPD2 (characterized quantum efficiency $\eta_{QE,2} = 90\%$), because we needed to reduce the bias current to prevent the detector from latching³⁸. This latching is probably caused by a nearby heater of the dilution refrigerator. It also results in a slow drift in the quantum efficiency of SNSPD1. The deduced $\eta_{path,i}$ come from the various optical elements in the beam path of the detection part and are in good agreement with their specified insertion losses.

Scattering rates and optomechanical coupling rate. From the total detection efficiency of resonantly generated cavity photons, we estimate the pair generation probability per write pulse (optical energy $E_{opt} \approx 40$ fJ) to be $p \approx 3.0\%$, which includes the effects of a finite starting temperature and leaked pump photons. The latter is calibrated by sending detuned optical pulses ($E_{opt} \approx 40$ fJ, $\omega_L = \omega_c - \omega_m - 2\pi \times 200$ MHz) to the device. The generated optomechanical sidebands are now blocked by the filters and only leaked pump photons are detected. We measure a suppression of the pump pulse by 84 dB compared to an on-resonance transmission. Thus, approximately 1 out of 25 photons detected during the write pulse is a leaked pump photon. Given the scattering rate and the energy

of the detuned pump pulse, we determine the single-photon coupling rate of our OMC to be

$$g_0 = \frac{\partial \omega_c}{\partial x} \sqrt{\frac{\hbar}{2m\omega_m}} = 2\pi \times 825 \text{ kHz}$$

With this coupling rate, we estimate the state-transfer efficiency of the red-detuned optical read-out pulse of $E_{\text{opt}} = 50 \text{ fJ}$ to be $\varepsilon_R = 3.7\%$, where $\hat{a}_{\text{opt,out}} \approx \sqrt{1 - \varepsilon_R} \hat{a}_{\text{opt,in}} + e^{i\phi} \sqrt{\varepsilon_R} \hat{a}_{\text{mech,in}}$. Here, $\hat{a}_{\text{opt,in(out)}}$ are the annihilation operators of the temporal optical input (output) mode of the cavity resonance, $\hat{a}_{\text{mech,in}}$ is the mechanical mode before the interaction and ϕ is an arbitrary, but fixed, phase between the inputs³⁹.

Definition and properties of the second-order correlation function. We define the normalized second-order correlation function for two, not necessarily different, modes α and β , and the respective annihilation operators \hat{a}_α and \hat{a}_β , to be (refs 3, 40 and references therein)

$$g_{\alpha\beta}^{(2)} = \frac{\langle : \hat{a}_\alpha^\dagger \hat{a}_\alpha \hat{a}_\beta^\dagger \hat{a}_\beta : \rangle}{\langle \hat{a}_\alpha^\dagger \hat{a}_\alpha \rangle \langle \hat{a}_\beta^\dagger \hat{a}_\beta \rangle}$$

where $: \hat{a} : \hat{a}^\dagger$ denotes normal ordering of the operators. For the autocorrelation of the optical field (photons scattered by the write pulse), $\alpha = \beta = o$, for the mechanical field $\alpha = \beta = m$ and for the cross-correlation $\alpha = o, \beta = m$. By introducing effective modes γ and δ it can be seen that this correlation function is independent of losses in the detection. Assuming the loss angles φ_α and φ_β for detection of modes α and β , respectively, we define the annihilation operators of the effectively detected modes γ and δ , $\hat{a}_{\gamma/\delta} = \cos(\varphi_{\alpha/\beta}) \hat{a}_{\alpha/\beta} + \sin(\varphi_{\alpha/\beta}) \hat{l}_{\alpha/\beta}$, by coupling the original modes α and β to modes l_α and l_β , represented by the annihilation operator $\hat{l}_{\alpha/\beta}$. As the detected modes γ, δ have frequencies in the optical domain, we assume the in-coupled modes l_α, l_β to be in their respective ground state. Tracing over l_α, l_β we find that $g_{\gamma\delta}^{(2)} = g_{\alpha\beta}^{(2)}$, that is, that the second-order correlation function is independent of losses or, in the case of the mechanical mode, of ‘ineffective’ partial state-transfer to the cavity mode. Thus, for example, $g_{mm}^{(2)}$ is equivalent to the autocorrelation of the photons scattered by the read pulse.

For autocorrelation measurements, we use a Hanbury Brown and Twiss set-up by splitting the mode on a symmetric beam-splitter and sending it to a pair of detectors. We define the modes detected by the individual detectors d_1, d_2 with their annihilation operators $\hat{a}_{1/2} = \cos(\theta) \hat{a}_\alpha \pm \sin(\theta) \hat{l}_d$, in which θ is the splitting angle of the beam-splitter and \hat{l}_d is the annihilation operator of the second input of the beam-splitter. The input state can, as before, be approximated to be in its vacuum state. We find that the autocorrelation of mode α is equal to the cross-correlation of the two detectors: $g_{\alpha\alpha}^{(2)} = g_{12}^{(2)}$. For a definition in terms of probabilities, see below.

Statistical analysis. No statistical methods were used to predetermine sample size.

Owing to the low detection probability, the uncertainty in the estimation of the second-order correlation functions is completely dominated by the estimation of the coincidence rate $\langle : \hat{a}_\alpha^\dagger \hat{a}_\alpha \hat{a}_\beta^\dagger \hat{a}_\beta : \rangle$ of the two modes α, β . As the absolute number of coincidences is low in some measurements, Gaussian statistics cannot be used for estimating uncertainties. Instead, we use the likelihood function based on the binomial distribution for estimating the probability p of the underlying process; that is, to obtain N counts in T tries $L(p, N, T) = p^N (1-p)^{T-N}/K$.

The normalization K is chosen such that $\int_0^1 L(p, N, T) dp = 1$. The upper and lower uncertainty σ_+ and σ_- are chosen numerically, such that they cover a 68% confidence interval around the maximum likelihood estimator $p_{\text{ML}} = N/T$: $\int_0^{p_{\text{ML}} - \sigma_-} L(p, N, T) dp = 0.16$, $\int_{p_{\text{ML}} + \sigma_+}^1 L(p, N, T) dp = 0.16$.

For the classical bound on the cross-correlation, $g_{\text{cb}} = \sqrt{g_{mm}^{(2)} g_{oo}^{(2)}}$, the likelihood functions of the individual autocorrelations are convoluted. Owing to their asymmetry, the maximum likelihood estimator of the classical bound is slightly lower than when using the individual maximum likelihood estimators: $g_{\text{cb,ML}} \leq \sqrt{g_{mm,ML}^{(2)} g_{oo,ML}^{(2)}}$.

As estimators for the cross-correlation function, the probabilities P of a coincident or single detection event during the read (R) and write pulse (W) were used, with $g_{\text{om}}^{(2)} = P(W \cap R) / [P(R)P(W)]$. This is valid for low event probabilities $P \ll 1$. Autocorrelations were estimated using probabilities of coincident and single detection events on individual SNSPDs (1, 2), $g_{yy}^{(2)} = P(X_1 \cap X_2) / [P(X_1)P(X_2)]$, during the evaluation periods of the write ($y = o, X = W$) and read ($y = m, X = R$) pulse, respectively. The statistics of the cross-correlation measurements are summarized in Extended Data Table 1.

For the read pulse, only the first 55 ns of the pulse were evaluated (see Fig. 3). A further reduction of the evaluation period to $t_{\text{eval}} = 30 \text{ ns}$ (R^*) has the advantage of reducing the influence of optical absorption of pump photons from the read pulse, while still obtaining solid statistics for the cross-correlation, $g_{\text{om}}^{(2)}(\Delta n = 0, \delta t = 100 \text{ ns}, t_{\text{eval}} = 30 \text{ ns}) = 19.6 - 2.8 + 3.9$. However, this reduction also results in a much lower state transfer efficiency $\varepsilon_R^* \approx 0.1\%$ (compared to $\varepsilon_R = 3.7\%$ above; see Methods section ‘Characterization of the detection scheme’). As a consequence we cannot obtain independent statistics on the autocorrelation function of the read pulse, because the number of pulse sequences is too low to observe coincidences during the reduced read pulse $N(R_1^* \cap R_2^*) = 0$. Thus, no independent classical bound g_{cb}^* can be obtained for this case. Because the measurement is identical to the one with a longer evaluation window in the first column of Extended Data Table 1, it is reasonable to assume the same autocorrelation of the mechanical state and, therefore, the same classical limit.

We note that slight differences in the polarization of the two input lasers and the optimal axis of the SNSPDs can lead to different detection rates of leaked pump photons between the read and the write pulse. Although this does not influence the cross-correlation measurement, it is important to use the same laser source for the sideband asymmetry measurements.

Interpretation of the cross-correlation measurements. Classical bound. The classical bound $g_{\text{cb,ML}}$ is found to be slightly greater than two, the value expected for a thermal state of the mechanical system (see Fig. 3c). Although this increase of the autocorrelation is not significant ($< 3\sigma$) in our measurements, a behaviour like this would be expected in the case of mixed thermal states of different temperatures, caused, for example, by fluctuations in the absorbed power. Effects that usually decrease the measured autocorrelation function of a thermal state, such as dark counts of the detectors and instantaneous heating by the read pulse, do not play a major role in our experiment, owing to the choice of pulse parameters. In conclusion, the classical bound in the present experiment is slightly elevated compared to cross-correlation experiments in atomic physics or nonlinear optics, where the classical bound is usually assumed to be⁴¹ less than two.

Decay of cross-correlations due to delayed heating. The cross-correlation can be interpreted as

$$g_{\text{om}}^{(2)} \approx \frac{\langle n_m \rangle_h}{\langle n_m \rangle}$$

in which $\langle n_m \rangle_h$ is the average number of mechanical excitations in the state heralded on a detection event of the write pulse (indicating the presence of an anti-Stokes scattered photon), and $\langle n_m \rangle$ is the average number of unheralded events (essentially probing the thermal excitation of the system when $p \ll 1$). In case of a delayed heating, the thermal occupation of the system is a function of the delay δt after the write pulse, $\langle n_{m,\text{th}} \rangle = \langle n_{m,\text{th}} \rangle / (\delta t)$. Assuming our cross-correlation is dominated by the thermal occupation, we obtain, for $\delta t \ll T_d$

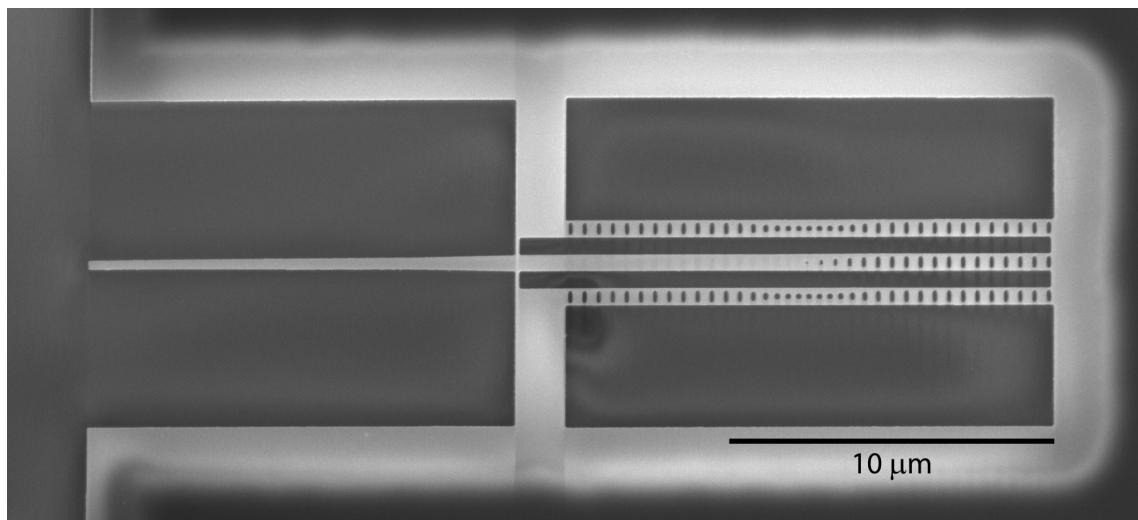
$$g_{\text{om}}^{(2)}(\delta t) \approx \frac{1 + \langle n_{m,\text{th}} \rangle / (\delta t)}{\langle n_{m,\text{th}} \rangle / (\delta t)}$$

which clearly decays in the case of substantial delayed heating, as is observed here (see Extended Data Fig. 3). Theoretical models of the complex thermodynamic non-equilibrium processes contain many device-dependent parameters¹⁷, which will be subject of further studies.

Estimation of the heralded single-phonon fidelity. In general, the toolbox of quantum optics provides a unique means for quantum state control of various systems⁴². As an example, we discuss the application of single-photon detection for the heralded generation of single-phonon Fock states of our mechanical resonator^{39,43,44}. To estimate the fidelity of the single-phonon state directly after heralding on the detection of a resonant photon generated by the write pulse, we need to know all the contributions to the diagonal of the density matrix, which are not a single phonon. These contributions can either be higher excitations by thermal contribution, multi-pair generation, or vacuum states by false-positive heralding events. Higher excitations can be estimated using the autocorrelation function of the heralded state, which is related to the cross-correlation function⁴⁵. Because we aim to estimate the state immediately after heralding it, we reduce the evaluation window of the read pulse as much as possible, while maintaining reasonable statistics on the cross-correlation (see Extended Data Table 1). From the measured $g_{\text{om}}^{(2)}(\Delta n = 0, \delta t = 100 \text{ ns}, t_{\text{eval}} = 30 \text{ ns}) = 19.6 - 2.8 + 3.9$, we infer an autocorrelation function for the heralded mechanical state of $g_{\text{mm,heralded}}^{(2)} \approx 0.22 \pm 0.04$, which approximately relates to the ratio of probabilities of higher excitations ($p_{n>1}$) to single phonon excitations ($p_{n=1}$): $2p_{n>1} \approx g_{\text{mm,heralded}}^{(2)} p_{n=1}^2$. In the meantime, the main contribution for non-zero $p_{n=0}$ (the probability of the heralded mechanical state being the ground state) is false-positive heralding events, that is, dark

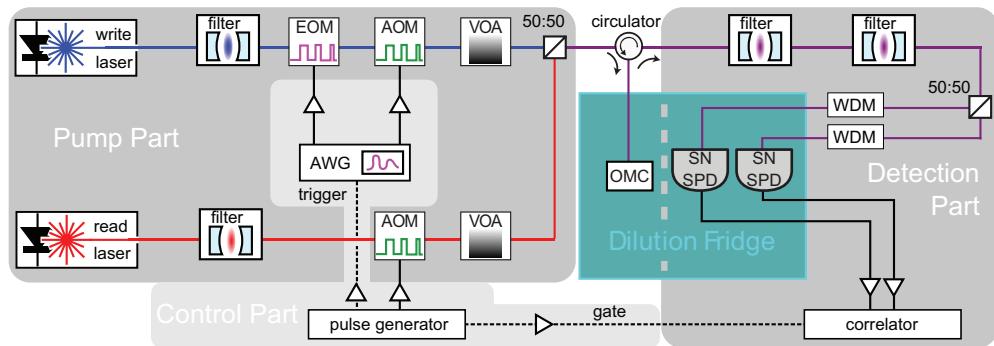
counts and leaked pump photons. With the known ratio of true-positive to false-positive heralding events (see Methods section 'Characterization of the detection scheme'), we obtain an estimate of $p_{n=0} \approx 1/25$. With these conservative estimates, we obtain a heralded Fock-state fidelity of $p_{n=1} = 87.7\% \pm 1.2\%$ on the basis of the standard system Hamiltonian of the optomechanical device³⁹.

31. Chan, J. *et al.* Optimized optomechanical crystal cavity with acoustic radiation shield. *Appl. Phys. Lett.* **101**, 081115 (2012).
32. Meenehan, S. M. *et al.* Silicon optomechanical crystal resonator at millikelvin temperatures. *Phys. Rev. A* **90**, 011803(R) (2014).
33. Fiore, V. *et al.* Storing optical information as a mechanical excitation in a silica optomechanical resonator. *Phys. Rev. Lett.* **107**, 133601 (2011).
34. Verhagen, E., Deléglise, S., Weis, S., Schliesser, A. & Kippenberg, T. J. Quantum-coherent coupling of a mechanical oscillator to an optical cavity mode. *Nature* **482**, 63–67 (2012).
35. Andrews, R. W. *et al.* Bidirectional and efficient conversion between microwave and optical light. *Nature Phys.* **10**, 321–326 (2014).
36. Bagci, T. *et al.* Optical detection of radio waves through a nanomechanical transducer. *Nature* **507**, 81–85 (2014).
37. Akram, U., Kiesel, N., Aspelmeyer, M. & Milburn, G. J. Single-photon optomechanics in the strong coupling regime. *New J. Phys.* **12**, 083030 (2010).
38. Natarajan, C. M., Tanner, M. G. & Hadfield, R. H. Superconducting nanowire single-photon detectors: physics and applications. *Supercond. Sci. Technol.* **25**, 063001 (2012).
39. Hofer, S. G., Wieczorek, W., Aspelmeyer, M. & Hammerer, K. Quantum entanglement and teleportation in pulsed cavity optomechanics. *Phys. Rev. A* **84**, 052327 (2011).
40. Mandel, L. & Wolf, E. *Optical Coherence and Quantum Optics* (Cambridge Univ. Press, 1995).
41. Sangouard, N., Simon, C., de Riedmatten, H. & Gisin, N. Quantum repeaters based on atomic ensembles and linear optics. *Rev. Mod. Phys.* **83**, 33–80 (2011).
42. Zoller, P., Cirac, J. I., Duan, L. & Garcia-Ripoll, J. J. in *Quantum Entanglement and Information Processing* (eds Estève, D., Raimond, J.-M. & Dalibard, J.) Session 79 of *Lecture Notes of the Les Houches Summer School 2003 Course 4* (Elsevier, 2004).
43. Vanner, M. R., Aspelmeyer, M. & Kim, M. S. Quantum state orthogonalization and a toolset for quantum optomechanical phonon control. *Phys. Rev. Lett.* **110**, 010504 (2013).
44. Galland, C., Sangouard, N., Piro, N., Gisin, N. & Kippenberg, T. J. Heralded single-phonon preparation, storage, and readout in cavity optomechanics. *Phys. Rev. Lett.* **112**, 143602 (2014).
45. Zhao, B. *et al.* A millisecond quantum memory for scalable quantum networks. *Nature Phys.* **5**, 95–99 (2009).

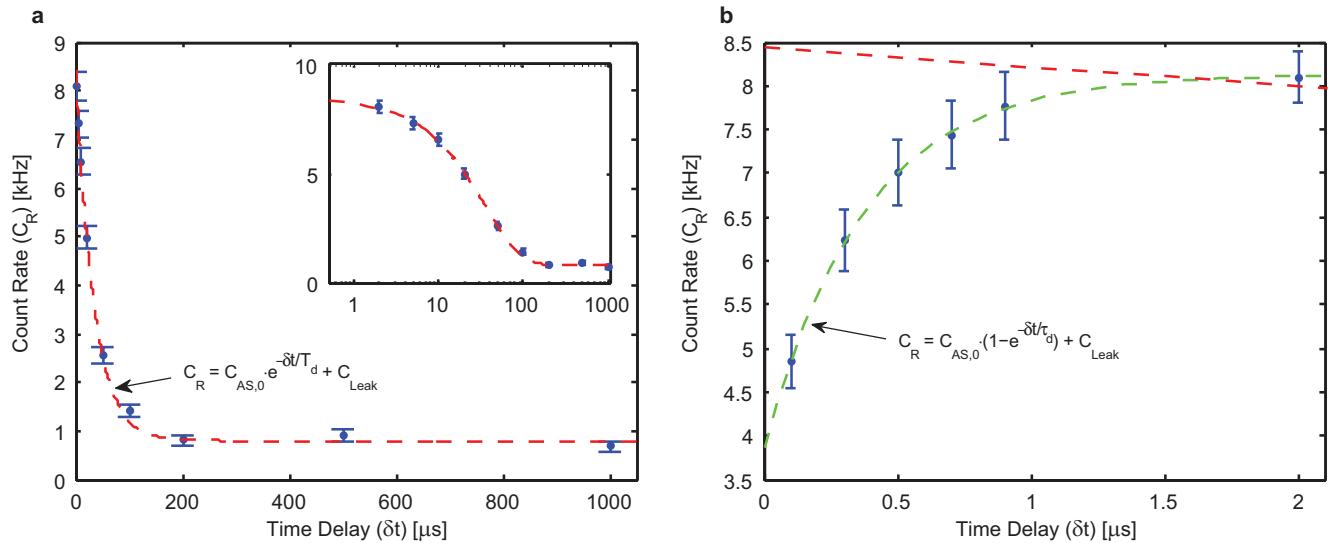


Extended Data Figure 1 | Optomechanical device. Shown is a scanning electron microscope image of a set of nanobeams, which are fabricated in silicon, as described in the text. Light is coupled into the central, adiabatically tapered waveguide through a lensed optical fibre (not shown)

from the left of the image. The field then evanescently couples to each nanobeam (top and bottom). The two devices have slightly different resonance frequency, which makes it possible to distinguish them.



Extended Data Figure 2 | Detailed experimental set-up. See Methods section ‘Set-up’ for a description.



Extended Data Figure 3 | Pump-probe measurement of the mechanical response. We send in a brief, intense blue-detuned optical pulse (pump) and measure the mechanical response via a red-detuned optical probe pulse as a function of pump-probe time delay (δt). **a**, Long-term mechanical response. The result fits well with a simple exponential decay (red dashed line; see equation in the plot) with a damping time constant (τ_d) of $34.4\text{ }\mu\text{s}$. The inset shows the same data/fit with a logarithmic scale on the x axis. $C_{AS,0}$ is the extrapolated C_{AS} ($\delta t = 0$). **b**, Short-term mechanical response. The data

are fitted to a simple exponential curve (green dashed line; see equation in the plot). The fitted time constant (τ_d) is $0.37\text{ }\mu\text{s}$. The fit curve of the long-term response (red dashed line) projected to $0\text{-}\mu\text{s}$ delay is also shown for comparison. Because the pump-pulse energies were five times stronger than those of the write pulses in the correlation experiment, it is expected that the delayed heating occurs on a longer timescale, owing to the temperature dependence of the thermal conductivity of silicon³². Error bars in **a** and **b** represent a 68% confidence interval.

Extended Data Table 1 | Counts of the cross-correlation measurements

δt	0.1 μ s	0.6 μ s	1.1 μ s	2.1 μ s	3.1 μ s	total	0.1 μ s*
$N(R \cap W)$	202	153	144	113	127		34
$N(R_1 \cap R_2)$	13	24	23	36	37		0
$N(W_1 \cap W_2)$	16	15	17	12	20	80	16
$N(R_1)$	13,172	16,523	18,751	18,316	23,629		966
$N(R_2)$	17,490	22,278	25,394	26,122	31,892		1145
$N(W_1)$	12,471	12,061	13,051	11,870	15,032	64,485	12,471
$N(W_2)$	19,409	18,616	20,176	18,329	23,601	100,131	19,409
T	38,806,017	38,829,923	39,958,216	35,712,159	47,964,927	201,271,242	38,806,017

The row labels $N(\text{event})$ represent the number of counts for a certain event, for example, detection of a photon during the measurement window of the read pulse on detector 1 or 2 ($R_{1/2}$), or the coincidence of a detection event of a subsequent write and read pulse on either detector 1 or 2 ($R \cap W = (R_1 \cup R_2) \cap (W_1 \cup W_2)$). T denotes the total number of pulse pairs sent to the optomechanical device. For the calculation of the autocorrelation function of the read pulse, only counts from the delay setting δt are used, because the delayed heating of the blue pulse (see Extended Data Fig. 3) influences the mechanical state. For the autocorrelation function of the write pulse, counts from all delay settings are summed, because the mechanical state is reinitialized by cryogenic cooling before measurement, independent of the delay δt . The numbers for this are summarized in the column labelled 'total'. The highest reported cross-correlation value was obtained by reducing the measurement window of the read pulse from 55 ns to 30 ns, with a delay of $\delta t = 100$ ns between the write and the read pulse. The counts for this evaluation window are presented in the column labelled '0.1 μ s*'. The underlying data set is the same as for the standard evaluation period of 55 ns, shown in the first column.

Uranium-mediated electrocatalytic dihydrogen production from water

Dominik P. Halter¹, Frank W. Heinemann¹, Julien Bachmann¹ & Karsten Meyer¹

Depleted uranium is a mildly radioactive waste product that is stockpiled worldwide. The chemical reactivity of uranium complexes is well documented, including the stoichiometric activation of small molecules of biological and industrial interest such as H₂O, CO₂, CO, or N₂ (refs 1–11), but catalytic transformations with actinides remain underexplored in comparison to transition-metal catalysis^{12–14}. For reduction of water to H₂, complexes of low-valent uranium show the highest potential, but are known to react violently and uncontrollably forming stable bridging oxo or uranyl species¹⁵. As a result, only a few oxidations of uranium with water have been reported so far; all stoichiometric^{2,3,16,17}. Catalytic H₂ production, however, requires the reductive recovery of the catalyst via a challenging cleavage of the uranium-bound oxygen-containing ligand. Here we report the electrocatalytic water reduction observed with a trisaryloxide U(III) complex $[(^{Ad,Me}ArO)_3mes]U$ (refs 18 and 19)—the first homogeneous uranium catalyst for H₂ production from H₂O. The catalytic cycle involves rare terminal U(IV)–OH and U(V)=O complexes, which have been isolated, characterized, and proven to be integral parts of the catalytic mechanism. The recognition of uranium compounds as potentially useful catalysts suggests new applications for such light actinides. The development of uranium-based catalysts provides new perspectives on nuclear waste management strategies, by suggesting that mildly radioactive depleted uranium—an abundant waste product of the nuclear power industry—could be a valuable resource.

To evaluate the electrocatalytic H₂O reduction with our U(III) complex, we performed reference experiments for uncatalysed H₂O reduction under identical conditions. As expected, the inert glassy carbon electrode possesses a considerably higher overpotential (η) for water reduction than does a platinum (Pt) electrode, as presented in the cyclic voltammograms in Fig. 1a (0.22 M H₂O in THF). We determined the thermodynamic potential for H₂ production under these conditions at -1.434 V versus Fc⁺/Fc using open-circuit potential (OCP) measurements (for details, see Supplementary Information). Cyclic voltammetry shows the onset potential for H₂ evolution at -2.25 V (versus Fc⁺/Fc, with $\eta = 0.82$ V) for a Pt electrode (red trace), whereas the onset potential is shifted to -3.25 V (versus Fc⁺/Fc, $\eta = 1.82$ V) for a glassy carbon electrode (black trace). The homogeneous uranium catalyst described and analysed in this study is the mesitylene (mes) anchored trisaryloxide uranium(III) complex $[(^{Ad,Me}ArO)_3mes]U$ (**1**), which incorporates bulky adamantyl (Ad) groups at the *ortho* positions and methyl (Me) groups at the *para* positions of the aryloxide arms (ArO). The addition of a minute amount of **1** (8.9×10^{-4} M, 0.4 mol%) to the electrolytic solution lowers the onset potential by 500 mV to -2.75 V (versus Fc⁺/Fc, $\eta = 1.32$ V) at the glassy carbon electrode, and substantially increases the current density from $-9.38 \mu A \text{ cm}^{-2}$ to $-0.232 \text{ mA cm}^{-2}$ (at -3.45 V versus Fc⁺/Fc, $\eta = 2.02$ V) at the glassy carbon electrode (blue trace).

A similar trend is observed in bulk electrolysis experiments, in which the catalyst **1** induces a substantial increase in the steady-state

current of the water reduction by a factor of ten, from $-17.6 \mu A$ to $-174 \mu A$ at a potential of -3.25 V (versus Fc⁺/Fc, $\eta = 1.82$ V). A series of bulk electrolyses at various potentials confirmed the catalytic activity of **1**, because the charge increases more rapidly at more negative potentials in the presence of **1** (Fig. 1b). The Faradaic yield of the H₂ evolution was determined to be close to 100% using a GC-TCD (gas chromatography with thermal conductivity detector; for details, see Supplementary Information). Control experiments using either $[UI_3(\text{THF})_4]$, a common precursor for trivalent uranium complexes, or pure ligand show no catalytic activity whatsoever (see Fig. 1b). In fact, the presence of uranium triiodide impedes H₂O reduction at very negative potentials. This proves that the catalytic activity is specifically due to compound **1**, and cannot be explained solely by the presence of a low-valent uranium ion. In addition, we excluded the formation of suspended uranium nanoparticles or an electroactive film on the electrode by multiple reference experiments, including electrolysis with added Hg (see Supplementary Information).

Electrochemical impedance spectroscopy (EIS) of the H₂O reduction allows for a quantification of the catalytic effect, because the charge transfer resistance determined by EIS is reduced by three orders of magnitude, from $1.23 \text{ M}\Omega$ to $3.70 \text{ k}\Omega$, at -3.25 V (versus Fc⁺/Fc), upon addition of **1** (Fig. 1c, d). A foot-of-the-wave analysis yields a turnover frequency of 10^6 h^{-1} at an overpotential of 1.3 V, the observed onset potential for our catalyst, whereas state-of-the-art homogeneous transition-metal H₂-evolving catalysts are reported with turnover frequencies ranging from 1 h^{-1} to $1.9 \times 10^7 \text{ h}^{-1}$ (ref. 20) for varying overpotentials. Important examples of electrocatalytic H₂ evolution include a nickel catalyst with the triflate salt of protonated dimethylformamid (DMF) in acetonitrile²¹, a cobalt catalyst with anilinium tetrafluoroborate in DMF²², a molybdenum oxo²³, or a cobalt-based catalyst in buffered aqueous phosphate solution²⁴. These diverse reagents illustrate the difficulties in comparing individual catalysts under different conditions. Comparison with other systems, where TOF values are not provided, can be made on the basis of overpotential decrease, Faradaic yields (usually between 85% and 100%), and an increase of the reductive current density at a suitable potential for uncatalysed H₂ evolution, which, for the majority of catalysts, leads to a 5- to 15-fold increase, but can reach up to 40-fold^{23,25–27}. The reported catalyst **1** reduces the overpotential by 0.5 V with respect to the glassy carbon electrode, has a Faradaic yield of 100%, a 25-fold reductive current increase at $\eta = 2$ V, and a turnover frequency of 10^6 h^{-1} at $\eta = 1.3$ V.

To provide a direct comparison with a recently published transition-metal system, we tested a Mo-based catalyst, $[(PY_5Me_2)MoO]^{2+}$, (provided by J. Long, C. Chang, and D. Zee) under the conditions used for our catalyst **1**. The current–potential curves (see Supplementary Information) of both systems are very similar.

To gain further insight into the catalytic reactions, and in particular to highlight the influence of bulky ligands for steric protection on the reactivity of U(III) complexes with H₂O in general, we carried out stoichiometric reactions.

¹Department of Chemistry and Pharmacy, Inorganic Chemistry, Friedrich-Alexander University Erlangen-Nürnberg (FAU), Egerlandstrasse 1, D-91058 Erlangen, Germany.

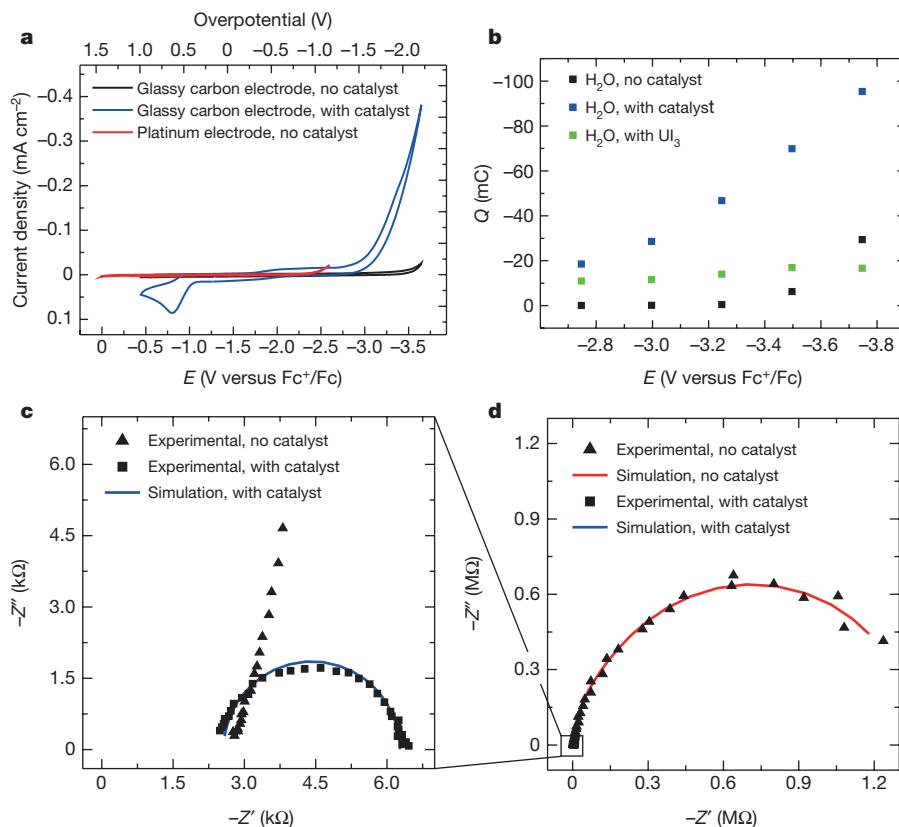


Figure 1 | Electrochemical characterization of catalyst 1. **a**, Cyclic voltammogram of the electrochemical H₂O reduction in THF with 0.1 M TBAPF₆ (20 μl H₂O in 5 ml THF, 0.22 M) on a glassy carbon electrode, without catalyst (black) and with 0.4 mol% catalyst 1 (blue). The onset potential for the water reduction is reduced by 0.5 V, and the reductive current density at the vertex potential increases from -0.027 mA cm⁻² to -0.382 mA cm⁻² after addition of the catalyst (5 mg, 0.4 mol%). For comparison, the H₂O reduction on a platinum electrode under similar conditions is shown (red). *E* is the potential, measured in volts. **b**, Plot of the charge *Q* passed during a 300-s electrolysis per run at different potentials *E* for uncatalysed H₂O electrolysis (black), in the presence of

0.4 mol% catalyst 1 (blue), and 0.4 mol% UI₃ (green), in THF with 0.1 M TBAPF₆. **c**, Close-up of the Nyquist plot in **d** for H₂O electrolysis at -3.25 V (versus Fc⁺/Fc) with and without catalyst 1, focusing on the catalysed reaction, in which *Z'* corresponds to the real part and *Z''* to the imaginary part of the impedance *Z*. Nyquist plots were simulated with the instrument's software to extract resistances and capacities (see Supplementary Information). **d**, Nyquist plot of the uncatalysed H₂O electrolysis at -3.25 V (versus Fc⁺/Fc). The charge transfer resistance is three orders of magnitude greater in the uncatalysed reaction than it is in the catalysed reaction, demonstrating the catalytic effect of compound 1.

Stoichiometric oxidation of the U(III) complex $[(^{(\text{Ad},\text{Me})}\text{ArO})_3\text{mes}\text{U}]$ (**1**) in THF or toluene with H₂O (0.05 M in THF) at room temperature results in the formation of the U(IV) complex $[(^{(\text{Ad},\text{Me})}\text{ArO})_3\text{mes}\text{U}(\text{OH})(\text{THF})]$ (**2-OH**) (Fig. 2a) and H₂, which was identified and quantified by GC-TCD experiments (see Supplementary Information). Single crystals of **2-OH** suitable for single-crystal XRD (X-ray diffraction) analysis were obtained as light-green needles at -35 °C. The tetravalent **2-OH** features a six-coordinate uranium ion in its solid-state molecular structure (Fig. 2b), with four oxygen atoms in the equatorial plane—the three aryloxide ligands with an average U-O_{Ar} bond of 2.188 Å, and a coordinated molecule of THF with a U-O_{THF} distance of 2.616(2) Å (the number in parentheses, here and elsewhere, is an estimate of the standard deviation derived from full-matrix least-squares refinement). The two axial positions are occupied by the hydroxyl ligand and the arene backbone of the chelator. This coordination geometry and bonding is similar to the analogous U(IV) halide complexes $[(^{(\text{Ad},\text{Me})}\text{ArO})_3\text{mes}\text{U}(\text{X})(\text{THF})]$ (with X = F, Cl, Br, I)²⁸. A U-arene_{centr.} distance of 2.703 Å and an average uranium-carbon bond length U-C_{Ar(av)} of 3.047 Å are in agreement with a uranium-arene δ-backbonding interaction.

The U-OH distance of the axially bound hydroxyl ligand to the uranium centre is 2.106(2) Å, which is in good agreement with those of comparable U(IV)-OH species^{3,16,17}. The hydroxide hydrogen position was derived from a difference Fourier synthesis and was treated using a riding model in the subsequent refinement cycles. Also, a hypothetical

terminal oxo ligand can be excluded, because U=O bonds are known to be much shorter, typically ranging from 1.818 Å to 1.859 Å (refs 29, 30). Infrared spectroscopy was used to confirm directly the identity of the U-OH group in **2-OH**. In a solid KBr matrix, complex **2-OH** exhibits two sharp O-H stretches with stretching frequencies, $\nu(\text{OH})$, centred at 3,659 cm⁻¹ and 3,630 cm⁻¹, probably due to the presence of two rotamers of the OH ligand (Fig. 2c). For an unambiguous assignment of the $\nu(\text{OH})$ stretches, the selectively deuterated complex $[(^{(\text{Ad},\text{Me})}\text{ArO})_3\text{mes}\text{U}(\text{OD})(\text{THF})]$ (**2-OD**) was synthesized from D₂O. As expected, the $\nu(\text{OD})$ bands were found at lower frequencies, 2,718 cm⁻¹ and 2,677 cm⁻¹, which is in accordance with the theoretically predicted bathochromic shifts of 990 cm⁻¹ based on a harmonic oscillator model.

The ¹H NMR spectrum of the paramagnetic U(IV) *f*² complex **2-OH** in C₆D₆ features the nine signals expected of the $(^{(\text{Ad},\text{Me})}\text{ArO})_3\text{mes}^{3-}$ ligand between 22.79 p.p.m. and -5.61 p.p.m., as well as two paramagnetically shifted signals of the coordinated THF (see Supplementary Information). The signal pattern is almost identical to the previously reported U(IV) fluoride complex $[(^{(\text{Ad},\text{Me})}\text{ArO})_3\text{mes}\text{U}(\text{F})(\text{THF})]$, for which a dynamic dissociation equilibrium of the THF ligand was shown to yield the observed pseudo *C*_{3v} symmetry²⁸. The OH proton was not detected, which is probably due to paramagnetic line broadening.

The visible–near-infrared electronic absorption spectrum of **2-OH** in THF displays eight absorption bands between 537 nm and

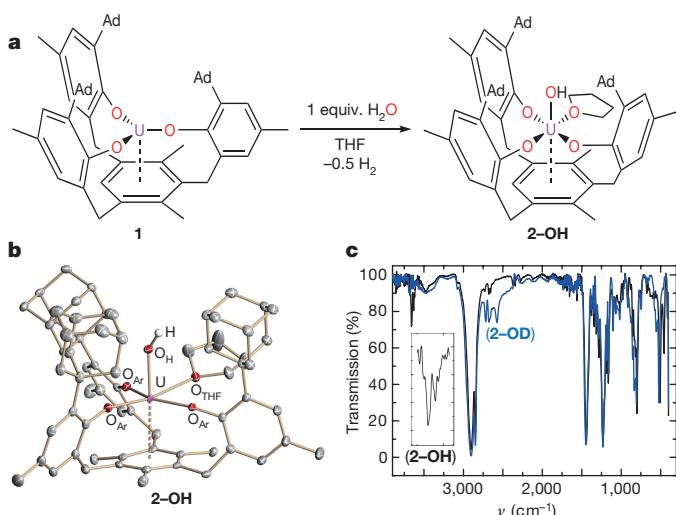


Figure 2 | Independent synthesis and characterization of the uranium(IV) hydroxo complex $[(^{(\text{Ad,MeArO})_3\text{mes}}\text{U}-\text{OH})]$ (2-OH). a, Synthesis of 2-OH with concomitant H_2 evolution. b, Molecular structure of the crystallographically characterized complex 2-OH in crystals of $\text{C}_{67}\text{H}_{84}\text{O}_5\text{U} \cdot 3(\text{C}_4\text{H}_8\text{O})$, with thermal ellipsoids at 50% probability. All hydrogen atoms except for the hydroxo H were omitted for clarity. c, Infrared vibrational spectra of 2-OH (black) and its isotopomer 2-OD (blue), showing the expected isotopic shift for the O-H stretching vibration ν . The inset is a close-up of the 2-OH spectrum, showing the two OH stretching frequencies at $\nu = 3,659 \text{ cm}^{-1}$ and $\nu = 3,630 \text{ cm}^{-1}$.

1,886 nm with extinction coefficients, ε , ranging from $61 \text{ mol}^{-1} \text{ cm}^{-1}$ to $231 \text{ mol}^{-1} \text{ cm}^{-1}$ (see Supplementary Information). The pattern and fine structure of the $f-f$ transition bands compare well with other U(IV) complexes of the $[(^{(\text{Ad,MeArO})_3\text{mes}}\text{U}(\text{X})(\text{THF})]$ system²⁸. Temperature-dependent SQUID (superconducting quantum interference device) experiments unambiguously characterize 2-OH as a U(IV) species with magnetic moments, μ_{eff} of $0.54 \mu_{\text{B}}$ at 2 K and $2.62 \mu_{\text{B}}$ at 300 K (in which μ_{B} is the Bohr magneton), as well as a temperature dependence reminiscent of the $[(^{(\text{Ad,MeArO})_3\text{mes}}\text{U}(\text{X})(\text{THF})]$ series mentioned above (see Supplementary Information)²⁸.

After identification of the reaction product 2-OH, we analysed the mechanism of its formation *in situ* using low-temperature X-band EPR (electron paramagnetic resonance) spectroscopy. This first mechanistic study of the fundamental reaction between H_2O and a molecular U(III) species provides basic chemical insight relevant to catalytic applications and nuclear waste treatment strategies. On the

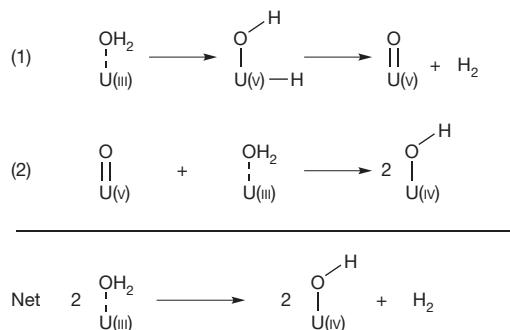


Figure 3 | Postulated mechanism for the reduction of H_2O by the U(III) complex 1, based on EPR results. The addition of H_2O to 1 probably yields a U(III) aquo species, which forms a fleeting U(v) hydroxo-hydrido intermediate, $[(^{(\text{Ad,MeArO})_3\text{mes}}\text{U}(\text{OH})(\text{H})]$, by intramolecular insertion; this hydroxo-hydrido species then decays to a U(v) oxo species by elimination of H_2 (reaction (1)). Subsequently, the U(IV) hydroxo complex 2-OH is formed in a comproportionation reaction between the U(v) oxo and the U(III) aquo species (reaction (2)). In the net reaction, two U(III) aquo complexes form two molecules of 2-OH and one equivalent H_2 .

basis of the results, a mechanism is postulated (Fig. 3) in which a U(III) aquo species undergoes an intramolecular insertion to form a U(v) hydroxo-hydrido intermediate. This intermediate then releases one equivalent of H_2 to form a U(v) oxo complex (Fig. 3, reaction 1). This proposed mechanism compares well to the reactivity reported for a molecular molybdenum-based catalyst for H_2O reduction, $[\text{PY}_5\text{Me}_2\text{MoO}]^{2+}$ (ref. 23, 31). In step 2 of the proposed mechanism, the U(v) oxo species comproportionates with a U(III) aquo complex to form two equivalents of the U(IV) hydroxo 2-OH (Fig. 3, reaction 2), a reaction that has been confirmed experimentally (see below and Supplementary Information for details).

To elucidate this mechanism, we performed time- and temperature-dependent EPR experiments with a reaction mixture of 1 and H_2O in a frozen toluene solution at 7.5 K (Fig. 4). Initially, a spectrum of the neat U(III) f^3 starting material (10 mM) in toluene was recorded, yielding an almost axial signal with g values centred at 1.56, 1.48, and 1.20 (see Supplementary Information), as expected for $[(^{(\text{Ad,MeArO})_3\text{mes}}\text{U}]$ (1)¹⁹. In the following measurement, a mixture of 1 (10 mM) in toluene with a sub-stoichiometric amount of H_2O (0.375 equiv.) was prepared. Under these dilute conditions the reaction takes about 2 h at room temperature for completion. Hence, the sample

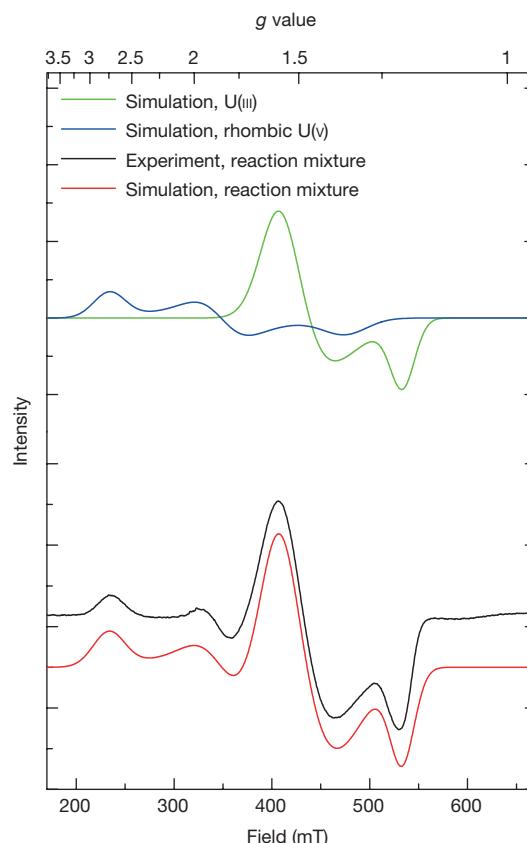


Figure 4 | X-band EPR spectrum of a frozen 10 mM toluene solution of 1 with a sub-stoichiometric amount of H_2O . The EPR data show a convoluted spectrum of two species: the U(III) starting material and a well-defined rhombic species, tentatively assigned to the fleeting U(v) hydroxo-hydrido species. Experimental conditions are as follows: temperature $T = 7.5 \text{ K}$, frequency $\nu = 8.96286 \text{ GHz}$, power $P = 1 \text{ mW}$, modulation width of 1.0 mT . The experimental spectrum (black) and simulation (red) under these conditions are shown. The best fit for the experimental spectrum is a convolution of the signal of 1 in toluene (simulated, green; g values at $g_1 = 1.56$, $g_2 = 1.48$, $g_3 = 1.20$, with line widths of $W_1 = 21.4 \text{ mT}$, $W_2 = 30.5 \text{ mT}$, $W_3 = 14.4 \text{ mT}$; relative weight of 1.0) and the signal of an additional, rhombic transient U(v) species (simulated, blue; g values at $g_1 = 2.73$, $g_2 = 1.83$, $g_3 = 1.35$, with line widths of $W_1 = 18.9 \text{ mT}$, $W_2 = 25.5 \text{ mT}$, $W_3 = 26.5 \text{ mT}$; relative weight of 0.70). The spectra are offset for ease of viewing.

was allowed to equilibrate for 5 min at room temperature and then flash-frozen in liquid nitrogen to trap potential intermediate species in a frozen solvent matrix. Indeed, we obtained a convoluted spectrum of at least two species: the U(III) starting material and another, well-defined rhombic species with simulated *g* values at 2.73, 1.83, and 1.35, consistent with an intermediate U(v) f^1 species (Fig. 4).

This rhombic U(v) species is tentatively assigned to a hydroxo-hydrido complex, formed by an intramolecular insertion reaction from an intermediate U(III) aquo species, as shown in Fig. 3. At 7.5 K, the U(v)-to-U(III) ratio increases with time until it reaches saturation after 10 min, which strongly suggests an intramolecular reaction. Starting from a U(III) aquo complex, this intramolecular reaction is likely to be the oxidative addition of the aquo ligand to yield the observed U(v) species that is the proposed U(v) hydroxo-hydrido complex. A hydroxo-hydrido complex would accommodate an additional ligand in the equatorial position, reminiscent of the labile coordinated THF in **2-OH**. Uranium complexes bearing the arene-anchored chelator $[(^{Ad,Me}ArO)_3mes]^3-$ do support the additional ligand in the equatorial plane, in contrast to other trisaryloxide chelators. This may be the reason why the closely related triazacyclononane derivative, $[(^{Ad,tBu}ArO)_3tacn]U$ (see Supplementary Information), does not show any catalytic activity for H_2O reduction. Furthermore, the electronic influence of the uranium-mesitylene δ -backbonding cannot be excluded; thus, identifying its importance is part of ongoing studies.

When the EPR sample is thawed, the U(v) hydroxo-hydrido complex that has accumulated decomposes thermally to form a U(v) oxo, which then comproportionates with the U(III) aquo complex to yield two equivalents of the EPR-silent U(iv) hydroxo species **2-OH** (Fig. 3, net reaction). Studies of the molecular and electronic structure of the U(v) oxo complex $[(^{Ad,Me}ArO)_3mes]U(O)$ are part of ongoing research; preliminary results yield an axially symmetric EPR spectrum, which deviates from the rhombic U(v) signal that is observed during the formation of the proposed, less-symmetric hydroxo-hydrido species. Hence, the U(v) hydroxo-hydrido complex is considered stable at 7.5 K in a frozen solvent matrix. Radical intermediates, such as H^\bullet or OH^\bullet , were not detected in the course of the reaction, but would have been visible prominently in EPR spectroscopy, had they been present. The role of the uranium oxo species as an intermediate during the catalytic cycle was further confirmed by the independent, stoichiometric synthesis of 2 equiv. of **2-OH** in a reaction of 1 equiv. of $[(^{Ad,Me}ArO)_3mes]U$ (**1**) with 1 equiv. of $[(^{Ad,Me}ArO)_3mes]U=O$, and 1 equiv. of H_2O , in quantitative yield (determined by 1H NMR; see Supplementary Information). After completion of the EPR reaction (2 h, at room temperature), the spectrum shows only signals of unreacted U(III) complex **1**, establishing the U(v) species as an intermediate during the formation of **2-OH**. The resultant U(iv) complex **2-OH** is EPR-silent, as is expected for a U(iv) complex with a $5f^2$ electron configuration and a non-magnetic singlet ground state (see above). To complete the analysis of the uranium-based electrocatalytic H_2O reduction, we investigated the stoichiometric regeneration of the U(III) species **1** from the U(iv) hydroxo complex **2-OH**. In addition to the chemical reduction of **2-OH** to **1** with KC_8 in 1H NMR experiments, square-wave and cyclic voltammetry revealed that an electrochemical reduction of **2-OH** by electrolysis is also possible.

Square-wave voltammetry experiments of **2-OH** reveal a broad reduction peak, ranging from -1.5 V to approximately -2.2 V (versus Fe^{+}/Fc). This redox feature is associated with the **2-OH**/ $2-OH^-$ couple, directly followed by the U(III)/U(II) redox transition at -2.5 V (versus Fe^{+}/Fc), recently published for **1** (see Supplementary Information)¹⁸. The characteristic, reversible U(III)/U(II) redox couple was also observed in cyclic-voltammetry experiments, in which a pure sample of **2-OH** was electrochemically reduced (using diffusion layer electrolysis at -2.2 V versus Fe^{+}/Fc), which confirmed the electrochemical conversion of $[(^{Ad,Me}ArO)_3mes]U(OH)(THF)$ (**2-OH**)

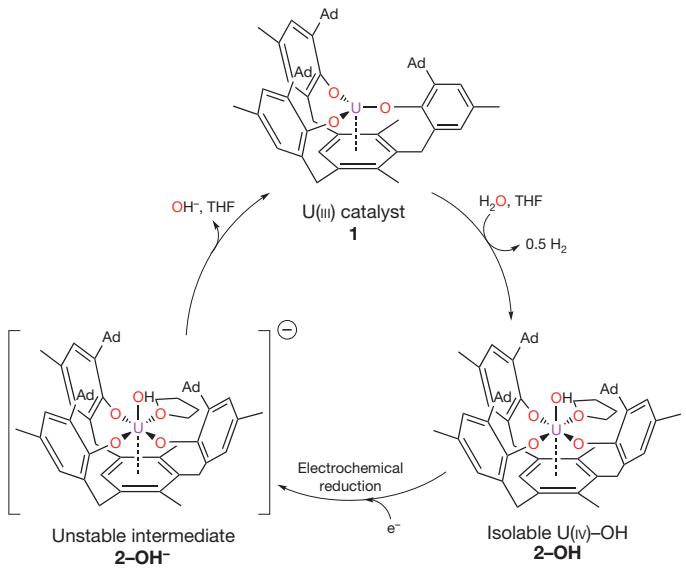


Figure 5 | Postulated electrocatalytic cycle for H_2 generation from H_2O in the presence of the homogeneous U(III) catalyst $[(^{Ad,Me}ArO)_3mes]U$ (1**). Step 1 (top to bottom-right), H_2 evolution and formation of $[(^{Ad,Me}ArO)_3mes]U(OH)(THF)$ (**2-OH**) through oxidation of **1** with H_2O . Step 2 (bottom-right to bottom-left), electrochemical reduction of **2-OH**, forming the transient anion **2-OH**⁻. Step 3 (bottom-left to top), elimination of OH^- from **2-OH**⁻ to regenerate catalyst **1**.**

to $[(^{Ad,Me}ArO)_3mes]U$ (**1**). Accordingly, the reaction involves the spontaneous elimination of OH^- (via U–O bond cleavage) from a reduced intermediary anionic complex $[(^{Ad,Me}ArO)_3mes]U(OH)(THF)]^-$ (**2-OH**⁻)—a reaction previously unknown for U(iv) complexes. This reductive splitting of the U–OH bond closes the proposed catalytic cycle for H_2 production from water.

In the first step of the proposed electrocatalytic cycle (Fig. 5), **2-OH** is formed in a chemical reaction by oxidation of **1** with H_2O and concomitant release of H_2 . Under the catalytic conditions (250 equiv. of H_2O) the formation of **2-OH** is effectively instantaneous, in contrast to the synthetic conditions (1 equiv. of H_2O), under which quantitative conversion of **1** to **2-OH** takes 2 h. In the second step, **2-OH** undergoes a one-electron electrochemical reduction to form the transient species **2-OH**⁻. The reduction potential of this step, observed approximately at -1.9 V (versus Fe^{+}/Fc) in square-wave-voltammetry experiments, does not correspond to an overpotential that is sufficient for direct proton reduction, but is associated with the **2-OH**/ $2-OH$ ⁻ redox couple, which is independent of the subsequent, purely thermal dihydrogen-forming step. Finally, in the third step, **2-OH**⁻ eliminates OH^- to regenerate **1** for another catalytic cycle. We further confirmed the presence of **2-OH** during the catalytic cycle by adding **2-OH** as the active form of the catalyst, instead of **1**, which similarly yielded the reported catalytic H_2O reduction.

The efficient electrocatalytic water reduction with a homogeneous uranium-based catalyst established here demonstrates the potential of actinides in catalysis. Future studies will be directed towards ligand derivatization for catalyst immobilization on electrode surfaces to further facilitate the (photo)electrocatalytic production of H_2 from water.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 21 October; accepted 7 December 2015.

Published online 25 January 2016.

1. Liddle, S. T. The renaissance of non-aqueous uranium chemistry. *Angew. Chem. Int. Ed.* **54**, 8604–8641 (2015).
2. Lukens, W. L., Beshouri, S. M., Blosch, L. L. & Andersen, R. A. Oxidative elimination of H_2 from $[Cp'_2U(\mu-OH)]_2$ to form $[Cp'_2U(\mu-O)]_2$, where Cp' is $1,3-(Me_3C)_2C_5H_3$ or $1,3-(Me_3Si)_2C_5H_3$. *J. Am. Chem. Soc.* **118**, 901–902 (1996).

3. John, G. H. *et al.* The synthesis, structural, and spectroscopic characterization of uranium(IV) perrhenate complexes. *Inorg. Chem.* **44**, 7606–7615 (2005).
4. Schmidt, A.-C., Heinemann, F. W., Lukens, W. W. & Meyer, K. Molecular and electronic structure of dinuclear uranium bis- μ -oxo complexes with diamond core structural motifs. *J. Am. Chem. Soc.* **136**, 11980–11993 (2014).
5. Cooper, O. *et al.* Multimetallic cooperativity in uranium-mediated CO_2 activation. *J. Am. Chem. Soc.* **136**, 6716–6723 (2014).
6. Summerscales, O. T., Cloke, F. G. N., Hitchcock, P. B., Green, J. C. & Hazari, N. Reductive cyclotramerization of CO to squareate by a $\text{U}(\text{III})$ complex: the X-ray crystal structure of $[(\text{U}(\eta\text{-C}_8\text{H}_6\text{Si}^{\text{i}}\text{Pr}_3\text{-1,4})_2)(\eta\text{-C}_5\text{Me}_4\text{H})_2](\mu\text{-}\eta^2\text{-C}_4\text{O}_4)$. *J. Am. Chem. Soc.* **128**, 9602–9603 (2006).
7. Frey, A. S., Cloke, F. G. N., Coles, M. P., Maron, L. & Davin, T. Facile conversion of CO/H_2 into methoxide at a uranium(III) center. *Angew. Chem. Int. Ed.* **50**, 6881–6883 (2011).
8. Odom, A. L., Arnold, P. L. & Cummins, C. C. Heterodinuclear uranium/molybdenum dinitrogen complexes. *J. Am. Chem. Soc.* **120**, 5836–5837 (1998).
9. Evans, W. J., Kozimor, S. A. & Ziller, J. W. A monometallic f element complex of dinitrogen: $(\text{Cs}^+\text{Me}_5)_3\text{U}(\eta^1\text{-N}_2)$. *J. Am. Chem. Soc.* **125**, 14264–14265 (2003).
10. Fox, A. R., Bart, S. C., Meyer, K. & Cummins, C. C. Towards uranium catalysts. *Nature* **455**, 341–349 (2008).
11. La Pierre, H. S. & Meyer, K. in *Progress in Inorganic Chemistry* Vol. 58 (ed. Karlin, K. D.) 303–415 (John Wiley & Sons, 2014).
12. Tway, C. L. Process for preparing a catalyst for the oxidation and ammonoxidation of olefins. US patent 6,916,763 (2005).
13. Wolf, A., Mleczko, L., Schlüter, O. F.-K. & Schubert, S. Integrated method for producing chlorine. US patent application 13/122,490 (2009).
14. Haber, F. & Le Rossignol, R. Process of making ammonia. US patent 999,025 (1911).
15. Korobkov, I. & Gambarotta, S. Trivalent uranium: a versatile species for molecular activation. *Prog. Inorg. Chem.* **54**, 321–348 (2005).
16. Natrajan, L., Mazzanti, M., Bezombes, J.-P. & Pécaut, J. Practical synthetic routes to solvates of $\text{U}(\text{OTf})_3$: X-ray crystal structure of $[\text{U}(\text{OTf})_3(\text{MeCN})_3]_n$, a unique $\text{U}(\text{III})$ coordination polymer. *Inorg. Chem.* **44**, 6115–6121 (2005).
17. Ariyaratne, K. A., Cramer, R. E., Jameson, G. B. & Gilje, J. W. Uranium-sulfilimine chemistry. Hydrolysis of $\text{Cp}^*\text{UCl}(\text{OH})(\text{HNSPh}_2)$, a metallocene terminal hydroxy complex of tetravalent uranium. *J. Organomet. Chem.* **689**, 2029–2032 (2004).
18. La Pierre, H. S., Kameo, H., Halter, D. P., Heinemann, F. W. & Meyer, K. Coordination and redox isomerization in the reduction of a uranium(III) monoarene complex. *Angew. Chem. Int. Ed.* **53**, 7154–7157 (2014).
19. La Pierre, H. S., Scheurer, A., Heinemann, F. W., Hieringer, W. & Meyer, K. Synthesis and characterization of a uranium(II) monoarene complex supported by δ backbonding. *Angew. Chem. Int. Ed.* **53**, 7158–7162 (2014).
20. Thoi, V. S., Sun, Y., Long, J. R. & Chang, C. J. Complexes of earth-abundant metals for catalytic electrochemical hydrogen generation under aqueous conditions. *Chem. Soc. Rev.* **42**, 2388–2400 (2013).
21. Helm, M. L., Stewart, M. P., Bullock, R. M., DuBois, M. R. & DuBois, D. L. A synthetic nickel electrocatalyst with a turnover frequency above 100,000 s^{-1} for H_2 production. *Science* **333**, 863–866 (2011).
22. Letko, C. S., Panetier, J. A., Head-Gordon, M. & Tilley, T. D. Mechanism of the electrocatalytic reduction of protons with diaryldithiophene cobalt complexes. *J. Am. Chem. Soc.* **136**, 9364–9376 (2014).
23. Karunadasa, H. I., Chang, C. J. & Long, J. R. A molecular molybdenum-oxo catalyst for generating hydrogen from water. *Nature* **464**, 1329–1333 (2010).
24. Cobo, S. *et al.* A Janus cobalt-based catalytic material for electro-splitting of water. *Nature Mater.* **11**, 802–807 (2012).
25. Cavell, A. C., Hartley, C. L., Liu, D., Tribble, C. S. & McNamara, W. R. Sulfinato iron(III) complex for electrocatalytic proton reduction. *Inorg. Chem.* **54**, 3325–3330 (2015).
26. Stubbert, B. D., Peters, J. C. & Gray, H. B. Rapid water reduction to H_2 catalyzed by a cobalt bis(iminopyridine) complex. *J. Am. Chem. Soc.* **133**, 18070–18073 (2011).
27. Das, A., Han, Z., Brennessel, W. W., Holland, P. L. & Eisenberg, R. Nickel complexes for robust light-driven and electrocatalytic hydrogen production from water. *ACS Catal.* **5**, 1397–1406 (2015).
28. Halter, D. P., La Pierre, H. S., Heinemann, F. W. & Meyer, K. Uranium(IV) halide (F^- , Cl^- , Br^- , and I^-) monoarene complexes. *Inorg. Chem.* **53**, 8418–8424 (2014).
29. Franke, S. M. *et al.* Uranium(III) complexes with bulky aryloxide ligands featuring metal-arene interactions and their reactivity towards nitrous oxide. *Inorg. Chem.* **52**, 10552–10558 (2013).
30. Arney, D. S. J. & Burns, C. J. Synthesis and structure of high-valent organouranium complexes containing terminal monooxo functional groups. *J. Am. Chem. Soc.* **115**, 9840–9841 (1993).
31. Sundstrom, E. J. *et al.* Computational and experimental study of the mechanism of hydrogen generation from water by a molecular molybdenum-oxo electrocatalyst. *J. Am. Chem. Soc.* **134**, 5233–5242 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank Y. Wu for support with the electrochemical impedance experiments. We also thank J. F. Berry and M. P. Bullock for discussions. For reference experiments, J. R. Long, C. J. Chang and D. Zee are acknowledged for the synthesis and donation of their Mo-based catalyst $[\text{PY}_5\text{Me}_2\text{MoO}](\text{B}(\text{C}_6\text{H}_3(\text{CF}_3)_2)_2)_2$ and D. J. Mindiola for the synthesis and crystallization of $[(\text{Ad}^{\text{tB}}\text{ArO})_3\text{tacn}]\text{U}(\text{OH})$. We acknowledge the Bundesministerium für Bildung und Forschung (BMBF, support codes 02NUK012C and 02NUK020C), the FAU Erlangen-Nürnberg, and COST Action CM1006 for financial support.

Author Contributions D.P.H., J.B., and K.M. planned the research and prepared the manuscript. D.P.H. performed the experiments. F.W.H. conducted the XRD analyses and refined structures. K.M. supervised the project in all aspects.

Author Information Atomic coordinates and structure factors for the reported crystal structures have been deposited in the Cambridge Crystallographic Data Centre under the accession code CCDC-1413741 (for **2-OH** from THF/n-pentane), CCDC-1401838 (for **2-OH** from THF), and CCDC-1437872 (for $[(\text{Ad}^{\text{tB}}\text{ArO})_3\text{tacn}]\text{U}(\text{OH})$). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.M. (karsten.meyer@fau.de).

METHODS

Materials. Compound **1**, $[(^{(\text{Ad},\text{Me})}\text{ArO})_3\text{mes}]\text{U}$, was prepared according to the previously described literature procedure¹⁸. Ultra-pure water was obtained from a Thermo Scientific TKA – Lab Type HP 4 purification system, and degassed before use. All other reagents were acquired from commercial sources and used as received. Solvents were purified using a two-column solid-state purification system (Glass Contour System), transferred to the glovebox without exposure to air, and stored over molecular sieves and sodium (where appropriate). NMR solvents were obtained packaged under argon and stored over activated molecular sieves and sodium (where appropriate) before use.

All air- and moisture-sensitive experiments were performed under dry nitrogen atmosphere using standard Schlenk techniques, or carried out in MBraun inert-gas gloveboxes containing an atmosphere of purified dinitrogen. The glovebox is equipped with a -35°C freezer. All glassware was dried by storage in an oven overnight (>8 h) at a temperature of $>160^\circ\text{C}$.

Synthesis of (2-OH) $[(^{(\text{Ad},\text{Me})}\text{ArO})_3\text{mes}]\text{U}(\text{OH})(\text{THF})$. A solution of $[(^{(\text{Ad},\text{Me})}\text{ArO})_3\text{mes}]\text{U}$ (30.0 mg, 0.0268 mmol) in THF (5 mL) was treated with 0.9 equiv. H_2O (0.05 M in THF, 490 μl , 0.0245 mmol) and stirred at room temperature for 2 h. The solvent was removed *in vacuo*, to yield an off-white powder, which was stirred in *n*-pentane (1 mL) for 1 h. The precipitate was filtered off over a frit, washed with *n*-pentane (3×1 mL), and dried *in vacuo*, to yield the title compound (**2-OH**) as an off-white powder in 87% yield (28.0 mg, 0.0232 mmol). Single crystals for XRD analysis were obtained from *n*-pentane diffusion into a concentrated solution of (**2-OH**) in THF at -35°C .

^1H NMR shifts (400 MHz, C_6D_6 , room temperature; br., broad; s, singlet; d, doublet): -5.61 p.p.m. (br. s, 18H), -3.07 p.p.m. (s, 9H), -2.38 p.p.m. (s, 3H), 0.04 p.p.m. (s, 3H), 1.57 p.p.m. (s, 3H), 10.81 p.p.m. (s, 6H), 11.06 p.p.m. (s, 9H), 11.95 p.p.m. (d, $J = 11$ Hz, 9H), 22.79 p.p.m. (br. s, 9H). Elemental analysis for $\text{C}_{67}\text{H}_{84}\text{O}_5\text{U}$: calculated C, 66.65; calculated H, 7.01; found C, 65.75; found H, 7.42. Infrared absorption bands (w, weak; m, medium; s, strong; vs, very strong): $3,659\text{ cm}^{-1}$ (m) O–H stretch, $3,630\text{ cm}^{-1}$ (m) O–H stretch, $2,903\text{ cm}^{-1}$ (vs), $2,847\text{ cm}^{-1}$ (vs), $1,440\text{ cm}^{-1}$ (vs), $1,375\text{ cm}^{-1}$ (w), $1,342\text{ cm}^{-1}$ (m), $1,315\text{ cm}^{-1}$ (m), $1,283\text{ cm}^{-1}$ (m), $1,232\text{ cm}^{-1}$ (vs), $1,209\text{ cm}^{-1}$ (s), $1,184\text{ cm}^{-1}$ (s), $1,163\text{ cm}^{-1}$ (s), $1,103\text{ cm}^{-1}$ (m), $1,072\text{ cm}^{-1}$ (w), $1,016\text{ cm}^{-1}$ (w), 980 cm^{-1} (w), 916 cm^{-1} (w), 854 cm^{-1} (m), 819 cm^{-1} (s), 800 cm^{-1} (vs), 750 cm^{-1} (m), 509 cm^{-1} (s), 457 cm^{-1} (s), 412 cm^{-1} (m).

Synthesis of (2-OD) $[(^{(\text{Ad},\text{Me})}\text{ArO})_3\text{mes}]\text{U}(\text{OD})(\text{THF})$. A solution of $[(^{(\text{Ad},\text{Me})}\text{ArO})_3\text{mes}]\text{U}$ (50.0 mg, 0.0447 mmol) in THF (5 mL) was treated with 1.0 equiv. D_2O (0.1 M in THF, 447 μl , 0.0447 mmol) and stirred at room temperature for 2 h. The solvent was removed *in vacuo*, to yield an off-white powder, which was stirred in *n*-pentane (1 mL) for 1 h. The precipitate was filtered off over a frit, washed with *n*-pentane (3×1 mL), and dried *in vacuo*, to yield the title compound (**2-OD**) as an off-white powder in 72% yield (39.0 mg, 0.0323 mmol).

^1H NMR shifts (400 MHz, C_6D_6 , room temperature): -5.39 p.p.m. (br. s, 18H), -2.96 p.p.m. (s, 9H), -2.15 p.p.m. (s, 3H), -0.03 p.p.m. (s, 3H), 1.30 p.p.m. (br. s, THF, 4H), 2.44 p.p.m. (s, 3H), 3.42 p.p.m. (br. s, THF, 4H), 10.58 p.p.m. (s, 6H), 10.93 p.p.m. (s, 9H), 11.80 p.p.m. (d, $J = 11.0$ Hz, 9H), 22.50 p.p.m. (br. s, 9H). Elemental analysis for $\text{C}_{67}\text{H}_{83}\text{DO}_5\text{U}$: calculated C, 66.65; calculated H, 7.01; found C, 66.71; found H, 7.20. Infrared absorption bands: $2,903\text{ cm}^{-1}$ (vs), $2,849\text{ cm}^{-1}$ (vs), $2,718\text{ cm}^{-1}$ (m) O–D stretch, $2,677\text{ cm}^{-1}$ (m) O–D stretch, $2,575\text{ cm}^{-1}$ (m), $1,445\text{ cm}^{-1}$ (vs), $1,375\text{ cm}^{-1}$ (w), $1,342\text{ cm}^{-1}$ (m), $1,315\text{ cm}^{-1}$ (m), $1,285\text{ cm}^{-1}$ (m), $1,233\text{ cm}^{-1}$ (vs), $1,184\text{ cm}^{-1}$ (s), $1,159\text{ cm}^{-1}$ (s), $1,103\text{ cm}^{-1}$ (m), $1,069\text{ cm}^{-1}$ (w), $1,018\text{ cm}^{-1}$ (m), 980 cm^{-1} (w), 916 cm^{-1} (w), 876 cm^{-1} (m), 854 cm^{-1} (s), 835 cm^{-1} (s), 806 cm^{-1} (s), 746 cm^{-1} (m), 555 cm^{-1} (m), 521 cm^{-1} (s), 413 cm^{-1} (m).

NMR spectroscopic measurements. ^1H NMR spectra were recorded on a JEOL ECX 400 at a probe temperature of 23°C . Chemical shifts δ are reported relative to residual ^1H resonances of the solvent in parts per million (ref. 32).

Electronic absorption spectroscopic measurements. Electronic absorption spectra were recorded from 200 nm to 2,500 nm (Shimadzu, UV-3600) in the indicated solvent at room temperature, and plotted from 440 nm to 2,150 nm to emphasize the *f*–*f* transitions.

Infrared vibrational spectroscopic measurements. Infrared spectra were recorded on a Shimadzu Affinity-1 CE FTIR instrument from 400 cm^{-1} to $4,000\text{ cm}^{-1}$. Solid samples of the compounds were homogenized with excess amounts of KBr and a pressed pellet was measured at room temperature.

EPR spectroscopic measurements. Variable-temperature X-band EPR spectra were recorded on a JEOL CW spectrometer, JES-FA200, equipped with an X-band Gunn diode oscillator bridge, a cylindrical mode cavity, and a helium cryostat. Spectra were recorded at $T = 7.5\text{ K}$ and simulated with the program W95EPR³³.

Elemental analysis. CH elemental microanalyses were obtained using Euro EA 3000 (Euro Vector) and EA 1108 (Carlo-Erba) elemental analysers in the Institute of Inorganic Chemistry at Friedrich-Alexander University Erlangen-Nürnberg.

Gas chromatographic measurements. All gas chromatography measurements were carried out on a Shimadzu GC-2010, equipped with a 5- \AA molecular sieves column, and a thermal conductivity detector (TCD), operated with N_2 as the carrier gas at 70°C . All samples (100 μl each) were injected with a gas-tight Hamilton syringe.

For the quantification of the H_2 content in the electrolysis cell head space, a calibration curve was prepared from multiple measurements with precisely adjusted H_2 concentrations in N_2 . The obtained calibration curve is shown in Supplementary Fig. 8. The Faradaic efficiency of the electrocatalytic cycle was studied by gas-chromatography-coupled electrolysis experiments, where the passed charge was compared to the amount of H_2 detected in GC-TCD experiments. The electrolysis cell was a standard Schlenk tube, sealed with a rubber septum, through which all electrodes were passed into the sample solution. After the electrolysis, 100 μl of the headspace were injected into the gas chromatograph with a gas-tight Hamilton syringe. Although great effort was made to seal the electrolysis cell, a H_2 leakage of 25% in 20 min was found in time-dependent GC-TCD experiments, resulting in an observed Faradaic yield significantly below 100%. To correct for the leakage, the same cell was equipped with a Pt working electrode to electrolyse an aqueous solution of 1.2 M NaOH for comparison. The electrolysis potential was adjusted such that the constant electrolysis current was identical to the electrolysis current of the electrocatalytic H_2O reduction using catalyst **1**. Indeed, matching H_2 contents were found after identical electrolysis time (75 min) for both the electrocatalytic cycle with **1**, and the platinum catalysed NaOH electrolysis, strongly suggesting identical Faradaic yields for both systems. Qualitative H_2 evolution in stoichiometric reactions of **1** with H_2O to form **2-OH** was also confirmed by GC-TCD experiments. Therefore, the synthesis was conducted in a closed reaction vessel, which was sealed with a septum to take a sample of the headspace.

Electrochemical experiments. Electrochemical experiments were carried out using a three-electrode set-up with a rotating glassy carbon working electrode (3 mm in diameter) and platinum rods as counter and reference electrodes. The potentiostat was a Metrohm μ Autolab Type-III. The entire set-up was placed inside a nitrogen-equipped glovebox, and only degassed, unstabilized, and dry THF (stored over activated molecular sieves) was used during the experiments. A platinized platinum electrode was freshly prepared before use according to previously described procedures³⁴, and handled in N_2 or H_2 atmosphere during the experiments. OCP measurements for the determination of the thermodynamic potential for H_2 formation, as introduced in ref. 35, were measured with a platinized platinum electrode, a platinum counter electrode, and a Ag wire reference electrode in a solution of 0.22 M H_2O in THF with 0.1 M TBAPF₆ under 1 atm of H_2 . EIS experiments were carried out using a three-electrode set-up with a glassy carbon working electrode, a platinum sheet counter electrode, and a Ag wire reference electrode. The potentiostat used for EIS and OCP measurements was a CompactStat potentiostat from Ivium Technologies and the impedance spectroscopy fits were performed by the instrument's software. All samples were measured in 0.1 M electrolyte solutions of TBAPF₆ (purchased from Sigma Aldrich and used without purification) in THF. Reported half-wave potentials were referenced to the Fc^+/Fc redox couple by adding recrystallized and purified ferrocene to the sample solution.

SQUID magnetization measurements. Magnetism data of crystalline powdered samples were recorded with a SQUID magnetometer (Quantum Design) at 10 kOe (2–300 K). Values of the magnetic susceptibility were corrected for the underlying diamagnetic increment ($\chi_{\text{dia}} = -694.04 \times 10^{-6}\text{ cm}^3\text{ mol}^{-1}$ for **2-OH**) by using tabulated Pascal constants and the effect of the blank sample holders (gelatin capsule/straw)³⁶. Samples used for magnetization measurements were checked for chemical composition and purity by CH elemental analysis and ^1H NMR spectroscopy.

Single-crystal X-ray crystallographic analyses. Suitable single crystals of each sample of **2-OH** as well as of $[(^{(\text{Ad},\text{tBu})}\text{ArO})_3\text{tacn}]\text{U}(\text{OH}) \cdot 2\text{Et}_2\text{O}$ were embedded in protective perfluoropolyalkylether oil and transferred to the cold nitrogen gas stream of the diffractometer. Intensity data for both samples of **2-OH** were collected using MoK_α radiation (wavelength $\lambda = 0.71073\text{ \AA}$) on a Bruker Kappa APEX 2 $I\mu\text{S}$ Duo diffractometer, equipped with QUAZAR focusing Montel optics; intensity data for $[(^{(\text{Ad},\text{tBu})}\text{ArO})_3\text{tacn}]\text{U}(\text{OH}) \cdot 2\text{Et}_2\text{O}$ were collected using MoK_α radiation ($\lambda = 0.71073\text{ \AA}$, graphite monochromator) on a Bruker-Nonius Kappa CCD diffractometer. Data were corrected for Lorentz and polarization effects; semi-empirical absorption corrections were performed on the basis of multiple scans using SADABS (version 2014/4 (2014) for **2-OH** from THF/*n*-pentane and **2-OH** from THF; version 2.06 (2002) for $[(^{(\text{Ad},\text{tBu})}\text{ArO})_3\text{tacn}]\text{U}(\text{OH}) \cdot 2\text{Et}_2\text{O}$; Bruker AXS, Inc.). The structures were solved by direct methods and refined by full-matrix least-squares procedures on F^2 using SHELXL-2014/6 (ref. 37). All non-hydrogen atoms were refined with anisotropic displacement parameters.

Complex **2–OH** from THF/*n*-pentane crystallized with a total of three molecules of THF per formula unit. All three THF molecules were disordered. Two alternative orientations were refined in each case resulting in site occupancies of 26.4(7)% and 73.6(7)% for the atoms O100–C104 and O110–C114, 43.6(10)% and 56.4(10)% for the atoms O200–C204 and O210–C214, and 49.1(9)% and 50.9(9)% for the atoms O300–C304 and O310–C314, respectively. *SAME*, *SIMU*, and *ISOR* restraints were applied in the refinement of the disorder. A tentative hydrogen position for the hydroxide H atom H5 could be determined from a difference Fourier synthesis. This hydrogen atom was treated using a riding model. All other hydrogen atoms were placed in positions of optimized geometry. The isotropic displacement parameters of all hydrogens were tied to those of their corresponding carrier atoms by a factor of 1.2 or 1.5.

In the crystal structure obtained for **2–OH** from THF, the ligand was heavily disordered. Two alternative orientations of two out of the three ligand arms were refined giving site occupancies of 57.9(3)% and 42.1(3)% for the affected atoms (atom names of the minor fraction were denoted with an additional A). One of the adamantyl groups was disordered with a different ratio of the two orientations giving site occupancies of 77.0(5)% and 23.0(5)% for the atoms C18–C25 and C18A–C25A, respectively. The compound crystallized with a total of 3.42 molecules of THF per formula unit. The partly occupied THF (O400–C404) is found in the place where the phenyl ring C29–C34 of the major fraction would be, but when the minor fraction of the ligand disorder is realized. *SAME*, *SIMU*, and *ISOR* restraints were applied. A tentative hydrogen position for the hydroxide H atom H5 could be determined from a difference Fourier synthesis, but, owing to both the heavy U atom and the strong disorder of the ligand, its positional parameters could not be refined. In the final refinement cycles, H5 was therefore placed in a position of optimized geometry and maximized electron density. All

other hydrogen atoms were also placed in positions of optimized geometry. The isotropic displacement parameters of all hydrogens were tied to those of their corresponding carrier atoms by a factor of 1.2 or 1.5.

In the crystal structure obtained for $[(^{(\text{Ad},\text{tBu})}\text{ArO})_3\text{tacn}]\text{U(OH)} \cdot 2\text{Et}_2\text{O}$, one of the *tBu* groups of the ligand was disordered. Two alternative orientations were refined giving site occupancies of 80(2)% and 20(2)% for the affected atoms C67–C69 and C67A–C69A, respectively. The compound crystallized with two molecules of diethyl ether per formula unit, one of which was disordered. Two alternative orientations were refined giving site occupancies of 57.9(9)% and 42.1(9)% for the atoms C201–C205 and C211–C215, respectively. *SAME*, *SIMU*, and, in part, *ISOR* restraints were used in the refinement of the disordered atoms. All hydrogen atoms including the hydroxide H were placed in positions of optimized geometry. The isotropic displacement parameters of all hydrogens were tied to those of their corresponding carrier atoms by a factor of 1.2 or 1.5.

32. Fulmer, G. R. *et al.* NMR chemical shifts of trace impurities: common laboratory solvents, organics, and gases in deuterated solvents relevant to the organometallic chemist. *Organometallics* **29**, 2176–2179 (2010).
33. Neese, F. *Electronic Structure and Spectroscopy of Novel Copper Chromophores in Biology*. Diploma thesis, Universität Konstanz (1993).
34. Feltham, A. M. & Spiro, M. Platinized platinum electrodes. *Chem. Rev.* **71**, 177–193 (1971).
35. Roberts, J. A. S. & Bullock, R. M. Direct determination of equilibrium potentials for hydrogen oxidation/production by open circuit potential measurements in acetonitrile. *Inorg. Chem.* **52**, 3823–3835 (2013).
36. Bain, G. A. & Berry, J. F. Diamagnetic corrections and Pascal's constants. *J. Chem. Educ.* **85**, 532–536 (2008).
37. Sheldrick, G. M. A short history of *SHELX*. *Acta Crystallogr. A* **64**, 112–122 (2008).

Ice stream activity scaled to ice sheet volume during Laurentide Ice Sheet deglaciation

C. R. Stokes¹, M. Margold^{1†}, C. D. Clark² & L. Tarasov³

The contribution of the Greenland and West Antarctic ice sheets to sea level has increased in recent decades, largely owing to the thinning and retreat of outlet glaciers and ice streams^{1–4}. This dynamic loss is a serious concern, with some modelling studies suggesting that the collapse of a major ice sheet could be imminent^{5,6} or potentially underway⁷ in West Antarctica, but others predicting a more limited response⁸. A major problem is that observations used to initialize and calibrate models typically span only a few decades, and, at the ice-sheet scale, it is unclear how the entire drainage network of ice streams evolves over longer timescales. This represents one of the largest sources of uncertainty when predicting the contributions of ice sheets to sea-level rise^{8–10}. A key question is whether ice streams might increase and sustain rates of mass loss over centuries or millennia, beyond those expected for a given ocean-climate forcing^{5–10}. Here we reconstruct the activity of 117 ice streams that operated at various times during deglaciation of the Laurentide Ice Sheet (from about 22,000 to 7,000 years ago) and show that as they activated and deactivated in different locations, their overall number decreased, they occupied a progressively smaller percentage of the ice sheet perimeter and their total discharge decreased. The underlying geology and topography clearly influenced ice stream activity, but—at the ice-sheet scale—their drainage network adjusted and was linked to changes in ice sheet volume. It is unclear whether these findings can be directly translated to modern ice sheets. However, contrary to the view that sees ice streams as unstable entities that can accelerate ice-sheet deglaciation, we conclude that ice streams exerted progressively less influence on ice sheet mass balance during the retreat of the Laurentide Ice Sheet.

Continental ice sheets are drained by a network of rapidly flowing ice streams with tributaries of intermediate velocity that extend far into their interiors¹¹. Towards the margins of modern ice sheets, many ice streams become confined within glacial troughs and are referred to as marine-terminating outlet glaciers, whereas others occupy regions of more subdued relief¹². Their large size (up to tens of kilometres wide and hundreds of kilometres long) and high velocity (up to thousands of metres per year) mean that they are an important mechanism through which ice is transferred to the ocean, and thereby affects the sea level. In contrast to climatically forced melting¹³, ice stream dynamics could introduce considerable nonlinearity into the response of ice sheets to external forcing^{5–7}. This potential nonlinearity is viewed as the major source of uncertainty in assessments of future changes in ice sheets and sea level^{1,6,8–10}, leading to the question of whether the drainage network of ice streams arises in response to climatically driven changes in ice sheet volume, or if it can evolve to drive changes beyond that which might be expected from climatic forcing alone.

Building on a recent inventory¹⁴ of 117 ice streams in the former North American Laurentide Ice Sheet (LIS; including the Innuitian Ice Sheet, but excluding the Cordilleran Ice Sheet) (Extended Data Fig. 1), we use the best-available ice margin chronology (based on about 4,000 dates; Extended Data Fig. 2)¹⁵ to ascertain the timing of their activity

during deglaciation (see Methods). Using the mapped extent of ice streams, we calculated the number of them operating over time and the percentage of the ice sheet perimeter that was streaming. We also explore their potential ice discharge during deglaciation, albeit with larger uncertainties, which we compare with output from numerical modelling of the ice sheet during deglaciation^{16,17} (see Methods).

When the LIS was at its maximum extent approximately 22,000 years ago (22 kyr ago), ice streams formed a drainage network with a velocity pattern resembling that of modern-day ice sheets (Fig. 1a, b). Early during deglaciation, numerous ice streams were located in major topographic troughs and drained the marine-based sectors of the northern and eastern margins of the ice sheet for several thousand years (Fig. 1b, c). Ice streaming along the land-terminating margins was more transient, and we find that numerous ice streams switched on and off in different locations during retreat. This empirical assessment of the duration of a large population of ice streams reveals that although some ice streams persisted for 5–10 kyr, about 40% operated for less than 2 kyr, with many (approximately 23%) operating for less than 0.5 kyr (Fig. 2).

Although ice streams activated and deactivated in different locations, we find no evidence for any episodes when the number of ice streams increased substantially (Fig. 3a). Throughout the period 22–15.5 kyr ago, there were about 50 ice streams, but this number subsequently dropped rapidly (for example, approximately 13 kyr ago and 11.5 kyr ago), with fewer than 10 ice streams operating after 11.5 kyr. When normalized by ice sheet volume, the number of ice streams is remarkably stable (Fig. 3b), with approximately 2 ice streams per 1,000,000 km³ of ice sheet volume for almost 10 kyr. The collapse of the LIS into Hudson Bay after 8.5 kyr ago triggered a final flurry of ice stream activity, but in a very small ice sheet.

At its maximum extent, approximately 27% of the LIS margin was streaming (Fig. 3c), which is very similar to that found for present-day Antarctica (Fig. 1a). This value decreased to between 25% and 20% from 16 kyr ago to 13 kyr ago, but then rapidly dropped to approximately 5% about 11 kyr ago. Similarly, our order-of-magnitude estimates of Laurentide ice stream discharge show no obvious increases during deglaciation (Fig. 3d). Rather, this dynamic component of mass loss was relatively stable from about 22 kyr ago to 15 kyr ago (remaining at a value of about 1,500 km³ yr⁻¹), but then rapidly decreased to <400 km³ yr⁻¹ by about 11 kyr ago, and then to <100 km³ yr⁻¹ about 9 kyr ago. When normalized by ice sheet volume, dynamic mass loss was relatively stable during the period 22–15 kyr ago, but then dropped about 13 kyr ago (Fig. 3e), before increasing temporarily as the ice sheet collapsed around Hudson Bay.

A comparison with estimates of total ice stream discharge from a previously published numerical model of the North American Ice Sheet complex¹⁶ and inferences from surface mass balance modelling at specific time steps¹⁷ indicates that model-derived Laurentide ice stream discharges are typically higher and more variable (Fig. 3f). Nevertheless, both empirical and modelled estimates show a decrease in ice stream discharge from about 15 kyr ago. Moreover, we find a

¹Department of Geography, Durham University, Durham, UK. ²Department of Geography, University of Sheffield, Sheffield, UK. ³Department of Physics and Physical Oceanography, Memorial University, St John's, Newfoundland, Canada. [†]Present address: Department of Physical Geography, Stockholm University, Stockholm, Sweden.

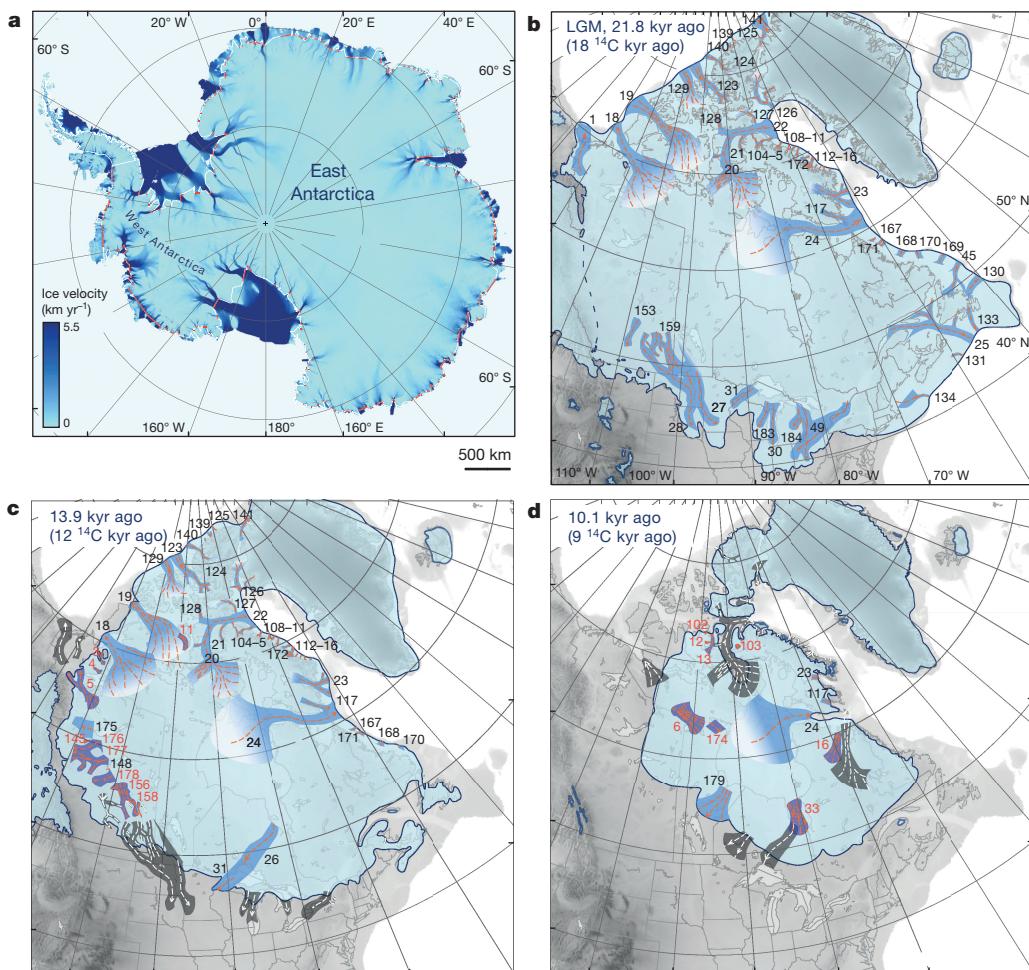


Figure 1 | Ice flow velocity of the Antarctic ice sheet compared (at the same spatial scale) with reconstructions of ice stream activity in the LIS at selected time steps. **a**, Present-day Antarctic ice sheet velocity¹¹, with red lines indicating where ice streams intersect the grounding line. **b–d**, Ice streams reconstructed for the LIS at the LGM (approximately 21.8 kyr ago) (**b**), 13.9 kyr ago (**c**) and 10.1 kyr ago (**d**), as labelled, with the radiocarbon (¹⁴C) dates given in parentheses. The extent of the LIS is indicated by the light blue shading and dark blue outline. The locations of

the Laurentide ice streams that were active at the given time are shown in blue and numbered in black, those that switched off within the preceding 1 kyr are shown in grey and those that switched on during the subsequent 1 kyr are shown in dark blue with numbers in red. Orange and white dashed lines with arrows show ice stream flow direction. The ice streams are labelled according to their inventory number¹⁴. The underlying topography is from GTOPO30 digital elevation data.

clear link between ice sheet volume and both the number of ice streams and the percentage of the ice sheet perimeter they occupied (Fig. 4a, b). A similar scaling is seen in both modelled and empirically derived discharge (Fig. 4c), but we acknowledge there are much larger uncertainties in our estimates of Laurentide ice stream discharge. The relative effect of ice streaming is seen more clearly by plotting ice sheet volume against the total ice stream discharge normalized by the ice sheet volume (Fig. 4d); these data indicate that the relative role of mass loss from streaming was unlikely to have increased as the ice volume decreased during deglaciation.

There are a number of factors that influence where ice streams develop, with previous work highlighting their strong association with topographic troughs, calving margins and soft sedimentary beds^{10,18,19}. Topography exerted a strong control on ice stream location in the LIS, particularly during early deglaciation (22–14 kyr ago) when its flow was steered by the major marine channels of the Canadian Arctic Archipelago and the high-relief coasts along the eastern margin. There is no glacial geomorphological evidence¹⁴ that these ice streams continued to operate once the ice sheet lost its marine margin and retreated onto lower-relief terrain. Thus, topographic troughs and the marine margin clearly modulated the number of ice streams operating over time (Figs 1 and 3a).

Elsewhere, ice streams were abundant on low-relief areas that were underlain by soft sedimentary bedrock and thick sequences of till. This includes the western and southern margins of the ice sheet^{14,19,20}, where numerous ice streams switched on and off during deglaciation, with marked changes in trajectory²⁰ (Fig. 1). These networks of sinuous ice streams deactivated as the ice margin withdrew onto the harder igneous and metamorphic rocks of the Canadian Shield, pointing to a geological control that explains the marked reduction in the number of ice streams from about 12 kyr ago (Fig. 3a, Extended Data Fig. 3). However, ice streams continued to activate over the low-relief crystalline bedrock of the Canadian Shield, with several large, wide (100–200 km) ice streams operating for very short periods (a few hundred years) during the final stages of deglaciation (beginning about 10 kyr ago; Fig. 1d; Extended Data Fig. 3)^{10,21}.

Although topography and underlying geology exerted an important influence on ice stream activity, we find no evidence for major ice sheet instabilities linked to ice stream activity that is reflected in the spatial re-organization of their drainage network (for example, marked increases in the number of ice streams, or individual ice streams widening or enlarging during ice sheet retreat). Rather, we find that the overall number of ice streams decreased and they occupied a progressively smaller percentage of the ice sheet perimeter. This finding implies that

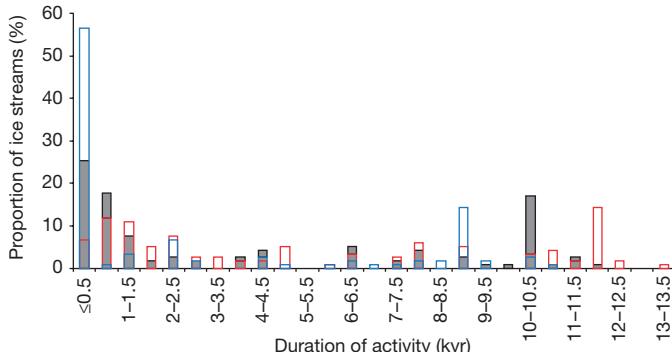


Figure 2 | Duration of individual ice streams in the LIS. Grey-filled bars represent the best estimate of the duration of each ice stream; blue and red outlines represent data assuming the minimum and maximum duration, respectively, for all ice streams (see Extended Data Fig. 1 for ice stream locations).

the final 4–5 kyr of deglaciation (beginning approximately 12 kyr ago) was largely driven by surface melt, which is corroborated by independent modelling of the surface mass balance of the ice sheet^{17,22} and inferences based on the density of subglacial meltwater channels (eskers)²³. Specifically, modelling of the surface energy balance of the ice sheet¹⁷ suggests that a transition from a positive to a negative surface mass balance occurred between 11.5 kyr ago and 9 kyr ago, when much of the LIS retreat occurred at rates two to five times faster than before 11.5 kyr ago. In that study¹⁷, volume losses not attributable to surface melting were assumed to be from dynamic discharge and, in broad agreement with our results (Fig. 3f), the modelling implies that dynamic discharge

decreased from about 15.5 kyr ago. Our range of discharge estimates at the Last Glacial Maximum (LGM) (750–2,300 km³) and in the early Holocene (100–700 km³ at 9 kyr ago) also fall within their inferred ranges¹⁷ (770–2,750 km³ and 0–1,650 km³, respectively). The major difference is that this previous surface energy balance modelling¹⁷ suggests that dynamic discharge increased from the LGM to a maximum (4,290–4,620 km³) around the time of Heinrich event 1 (H1; 15.5 kyr ago) when the modelled surface mass balance is largest. A positive mass balance is temporarily induced in the other ice sheet modelling¹⁶ shown in Fig. 3f to facilitate a large dynamic discharge from the Hudson Strait ice stream during H1. We do not depict such extreme discharge at this time because our approach is based on modern ice stream data that are unlikely to capture such extreme discharges. However, we find no obvious spatial reorganization²⁴ of ice streams during or immediately after H1. This finding suggests that H1 had a limited effect on the wider ice sheet drainage network, and points to extreme velocity fluctuations for specific ice streams (for example, Hudson Strait)²⁵, which we are unable to constrain, and/or mechanisms that do not invoke major collapses of ice sheets, such as ice shelf break-up²⁶. By contrast, we note some reorganization of ice streaming following, but not before or during, Meltwater Pulse 1A (that began about 14.6 kyr ago)²⁷. The saddle collapse that occurred during separation of the Laurentide and Cordilleran ice sheets has been hypothesized to have contributed to this event²⁸ and we observe several short-lived ice streams in this region after the collapse, but with a concomitant decrease in ice stream activity along the southern margin (Fig. 1c).

It is important to consider whether ice streaming in the LIS offers an analogue for modern-day ice sheets. Although the ocean–climate forcing would have been different during deglaciation of the LIS, there is no empirical evidence or theoretical reasoning to suppose that Laurentide

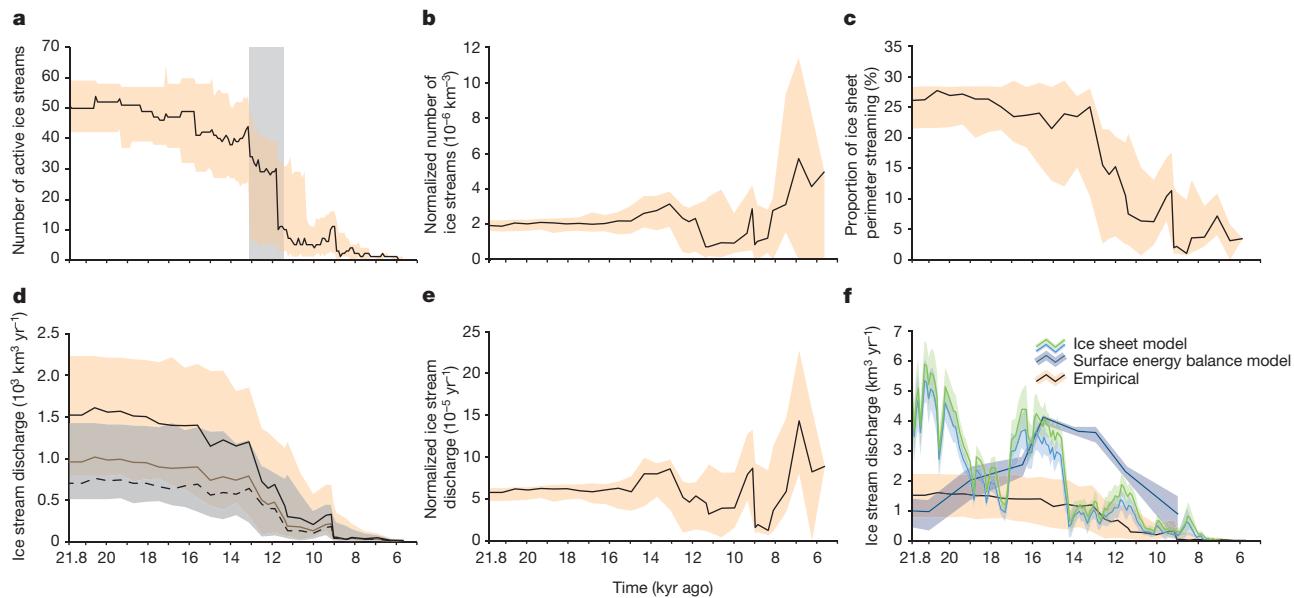


Figure 3 | Ice stream activity during deglaciation of the LIS. **a**, Number of active ice streams during deglaciation. The black line is the best estimate and the orange shading indicates the uncertainty in the age-bracketing of ice streams. The grey vertical bar shows the time when the ice sheet margin transitioned from a predominantly soft to a predominantly hard bed. Data are measured in intervals of 100 yr. **b**, Number of active ice streams normalized by LIS volume obtained from data-calibrated numerical modelling¹⁶. Data are measured in intervals of 500 yr; line and shading as in **a**. **c**, Percentage of the ice sheet perimeter that was streaming. Data are measured in intervals of 500 yr; line and shading as in **a**. **d**, First-order estimate of the total Laurentide ice stream discharge based on a width–discharge regression from modern-day ice stream data (black line; see Methods). Orange shading indicates the uncertainty in both the age-bracketing of the ice streams and the discharge uncertainty from the 95% confidence intervals of the regression. For comparison,

brown line and grey shading shows the same range of discharges obtained from the regression without two obvious outliers (Extended Data Fig. 10). Dashed line shows discharge from a cruder, two-state approximation for the best estimate of ice stream duration (see Methods). **e**, Ice stream discharge (black line in **d**) normalized by LIS volume from numerical modelling¹⁶; orange shading as in **a**. **f**, Empirical ice discharge (black line and orange shading in **d**) compared to ice stream discharge generated from the mean of a data-calibrated numerical modelling ensemble of the LIS¹⁶. The light blue and green lines show ice stream discharge from grid cells on the margin of the grounded ice in which the modelled velocities were $>500 \text{ m yr}^{-1}$ and $>100 \text{ m yr}^{-1}$, respectively (both with 1σ uncertainties). The dark blue line shows the discharge that is inferred from previous modelling of the surface mass balance of the ice sheet¹⁷, with shading indicating uncertainty as propagated from surface mass balance results (see ref. 17).

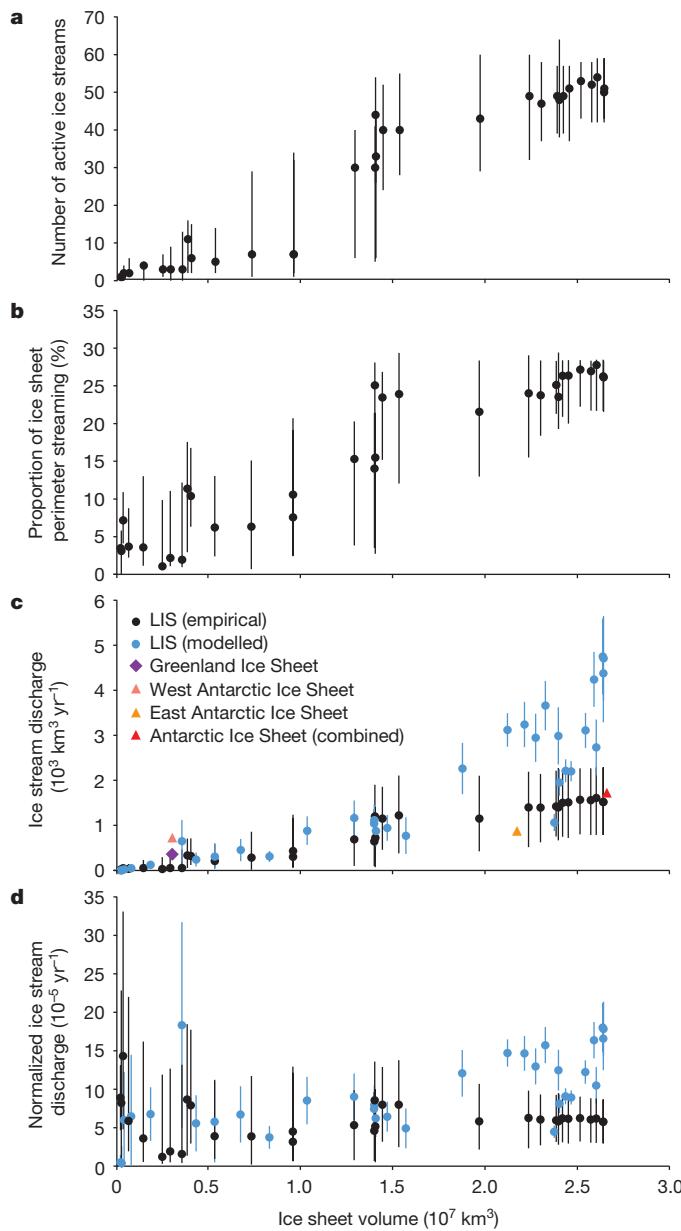


Figure 4 | Indicators of Laurentide ice stream activity plotted against ice sheet volume. **a**, Number of active ice streams. Black dots show results using best estimates of ice stream duration with vertical lines indicating the uncertainty in the age-bracketing. **b**, Percentage of the ice sheet perimeter that was streaming (symbols as in **a**). **c**, Total ice stream discharge as determined from our empirical calculations (black dots) show best estimate and vertical lines indicate both the age and discharge uncertainties from Fig. 3d) and from numerical modelling¹⁶ (blue dots) show mean and vertical lines show 2σ uncertainties. Modelled discharge is extracted from grid cells on the margin of the grounded ice in which the modelled velocities were $>500 \text{ m yr}^{-1}$ (blue line in Fig. 3f). Data for the present-day ice sheets in Greenland (purple diamond), West Antarctica (pink triangle), East Antarctica (orange triangle) and West and East Antarctica combined (red triangle) are also plotted using recent ice stream discharge^{2,3} and ice sheet volume^{29,30} estimates. **d**, Total ice stream discharge as determined from modelled and empirical estimates in **c** normalized by ice sheet volume. In all panels, ice sheet volume is derived from the mean of a best-performing ensemble of previously published data-calibrated numerical models¹⁶ (see Methods). The low modelled streaming fraction at volumes of about $2.4 \times 10^7 \text{ km}^3$ (**c**, **d**) are due to the dynamic facilitation of H1 in that modelling (see Methods). Empirical discharges for small ice volumes ($<0.2 \times 10^7 \text{ km}^3$) have highly under-represented uncertainties, owing to higher uncertainties in ice sheet volume close to final deglaciation.

ice streams should behave in a fundamentally different manner. Our reconstructed pattern of ice streams at the LGM is remarkably similar to the velocity pattern of the Greenland and Antarctic ice sheets, and we note that Laurentide ice streams drained a similar proportion of the ice sheet perimeter when it was a similar size to present-day Antarctica (Fig. 1a, b). Large sectors of the LIS occupied similar physiography to modern-day ice sheets, with ice streams exhibiting a similar size, shape and spatial organization along its marine margins. The most obvious difference is that the LIS retreated onto a low-relief, hard bedrock terrain and had ice streams that terminated on land and produced large, low-relief lobes along much of the southern and western margins^{19,20}. Although these have been likened to some West Antarctic ice streams¹⁹, they have no modern-day analogue. However, despite the fact that all modern-day ice streams are marine-terminating, large parts of the Greenland and East Antarctic ice sheets will have land-based margins if they continue to deglaciate^{29,30}, which might be within a few millennia in Greenland¹. Our analysis confirms that the geology and topography over which modern-day ice sheets retreat will be a key determinant of where ice streams are likely to activate and deactivate⁸. However, we also find a strong dependence between ice sheet volume and ice stream activity that also holds for modern-day ice sheets (Fig. 4c), and which hints at a more regulatory role in ice sheet dynamics than was previously recognized. This result does not preclude instabilities on decadal to centennial timescales^{5–7}, but suggests that if modern-day ice sheets continue to deglaciate, then ice streams are likely to switch off, and their relative contribution to mass loss may decrease over several millennia, with final deglaciation accomplished most effectively by surface melt^{17,22}.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 July; accepted 16 December 2015.

1. Alley, R. B., Clark, P. U., Huybrechts, P. & Joughin, I. Ice-sheet and sea-level changes. *Science* **310**, 456–460 (2005).
2. Rignot, E. & Kanagaratnam, P. Changes in the velocity structure of the Greenland Ice Sheet. *Science* **311**, 986–990 (2006).
3. Rignot, H. D. *et al.* Recent Antarctic ice mass loss from radar interferometry and regional climate modelling. *Nature Geosci.* **1**, 106–110 (2008).
4. Rignot, E., Velicogna, I., van den Broeke, M. R., Monaghan, A. & Lenaerts, J. T. M. Acceleration of the contribution of the Greenland and Antarctic ice sheets to sea level rise. *Geophys. Res. Lett.* **38**, L05503 (2011).
5. Golledge, N. R. *et al.* The multi-millennial Antarctic commitment to future sea-level rise. *Nature* **526**, 421–425 (2015).
6. Feldmann, J. & Levermann, A. Collapse of the West Antarctic Ice Sheet after local destabilization of the Amundsen Basin. *Proc. Natl Acad. Sci. USA* **112**, 14191–14196 (2015).
7. Joughin, I., Smith, B. E. & Medley, B. Marine ice sheet collapse potentially under way for the Thwaites Glacier Basin, West Antarctica. *Science* **344**, 735–738 (2014).
8. Ritz, C. *et al.* Potential sea-level rise from Antarctic ice-sheet instability constrained by observations. *Nature* **528**, 115–118 (2015).
9. Intergovernmental Panel on Climate Change (IPCC) in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. *et al.*) 3–29 (Cambridge Univ. Press, 2013).
10. Kleman, J. & Applegate, P. J. Durations and propagation patterns of ice sheet instability events. *Quat. Sci. Rev.* **92**, 32–39 (2014).
11. Rignot, E., Mouginot, J. & Scheuchl, B. Ice flow of the Antarctic ice sheet. *Science* **333**, 1427–1430 (2011).
12. Rose, K. E. Characteristics of ice flow in Marie Byrd Land, Antarctica. *J. Glaciol.* **24**, 63–75 (1979).
13. Nghiem, S. V. *et al.* The extreme melt across the Greenland ice sheet in 2012. *Geophys. Res. Lett.* **39**, L20502 (2012).
14. Margold, M., Stokes, C. R., Clark, C. D. & Kleman, J. Ice streams in the Laurentide Ice Sheet: a new mapping inventory. *J. Maps* **11**, 380–395 (2015).
15. Dyke, A. S., Moore, A. & Robertson, L. *Deglaciation of North America. Open File Report No. 1574* (Geological Survey of Canada, 2003).
16. Tarasov, L., Dyke, A. S., Neal, R. M. & Peltier, W. R. A data-calibrated distribution of deglacial chronologies for the North American ice complex from glaciological modeling. *Earth Planet. Sci. Lett.* **315–316**, 30–40 (2012).
17. Ullman, D. J., Carlson, A. E., Anslow, F. S., LeGrande, A. N. & Licciardi, J. M. Laurentide ice-sheet instability during the last deglaciation. *Nature Geosci.* **8**, 534–537 (2015).

18. Winsborrow, M. C. M., Clark, C. D. & Stokes, C. R. What controls the location of ice streams? *Earth Sci. Rev.* **103**, 45–59 (2010).
19. Clark, P. U. Surface form of the southern Laurentide Ice Sheet and its implications to ice-sheet dynamics. *Geol. Soc. Am. Bull.* **104**, 595–605 (1992).
20. Ó Cofaigh, C., Evans, D. J. A. & Smith, I. R. Large-scale reorganisation and sedimentation of terrestrial ice streams during late Wisconsinan Laurentide Ice Sheet deglaciation. *Geol. Soc. Am. Bull.* **122**, 743–756 (2010).
21. Stokes, C. R. & Clark, C. D. Laurentide ice streaming over the Canadian Shield: a conflict with the soft-bedded ice stream paradigm? *Geology* **31**, 347–350 (2003).
22. Carlson, A. E. *et al.* Surface-melt driven Laurentide Ice Sheet retreat during the early Holocene. *Geophys. Res. Lett.* **36**, L24502 (2009).
23. Storrar, R. D., Stokes, C. R. & Evans, D. J. A. Increased channelization of subglacial drainage during deglaciation of the Laurentide Ice Sheet. *Geology* **42**, 239–242 (2014).
24. Mooers, H. D. & Lehr, J. D. Terrestrial record of Laurentide Ice Sheet reorganisation during Heinrich events. *Geology* **25**, 987–990 (1997).
25. Hemming, S. R. Heinrich events: massive late Pleistocene detritus layers of the North Atlantic and their global climate imprint. *Rev. Geophys.* **42**, RG1005 (2004).
26. Marcott, S. A. *et al.* Ice-shelf collapse from subsurface warming as a trigger for Heinrich events. *Proc. Natl Acad. Sci. USA* **108**, 13415–13419 (2011).
27. Deschamps, P. *et al.* Ice-sheet collapse and sea-level rise at the Bolling warming 14,600 years ago. *Nature* **483**, 559–564 (2012).
28. Gregoire, L., Payne, A. J. & Valdes, P. J. Deglacial rapid sea level rises caused by ice-sheet saddle collapses. *Nature* **487**, 219–222 (2012).
29. Bamber, J. L. *et al.* A new bed elevation dataset for Greenland. *The Cryosphere* **7**, 499–510 (2013).
30. Fretwell, P. *et al.* Bedmap 2: improved ice bed, surface and thickness datasets for Antarctica. *The Cryosphere* **7**, 375–393 (2013).

Acknowledgements This research was funded by a Natural Environment Research Council award NE/J00782X/1 (C.R.S. and C.D.C.). Landsat imagery and the GTOPO30 digital elevation model were provided free of charge by the US Geological Survey Earth Resources Observation Science Centre.

Author Contributions C.R.S. designed the study and wrote the proposal with C.D.C. M.M. generated the data on the timing of Laurentide ice streams, modern-ice stream discharge and Laurentide ice stream discharge, and produced the figures, with input from C.R.S. and C.D.C. L.T. contributed data from numerical modelling. All authors contributed to the analyses and interpretations of the data. C.R.S. wrote the manuscript with input from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.R.S. (c.r.stokes@durham.ac.uk).

METHODS

Identifying Laurentide ice stream locations. An ice stream is a region in a grounded ice sheet that flows much faster than the regions on either side³¹. Where the fast-flowing ice becomes bordered by exposed rock (for example, in high-relief fjord landscapes), they are usually referred to as marine-terminating outlet glaciers. These outlet glaciers typically initiate as ice streams and so we use the term 'ice stream' throughout, but include outlet glaciers.

We use a recently published inventory¹⁴ of ice streams in the Laurentide Ice Sheet (LIS), which includes 117 ice streams (Extended Data Fig. 1). These were identified on the basis of previously published evidence, complemented with new mapping using satellite imagery and Digital Elevation Models (DEMs) on land, and bathymetric data and swath bathymetry for submerged areas¹⁴. The systematic nature of the new mapping from across the entire ice sheet bed means it is very unlikely that any major ice streams have been missed¹⁴.

Ice streams are easily distinguishable on an ice sheet bed from a variety of evidence that is now well established in the literature^{14,20,21,32–43}. Their spatially discrete, enhanced flow creates a distinctive bedform imprint that is immediately recognizable and characterized by several geomorphological criteria³². These include highly elongated subglacial bedforms (mega-scale glacial lineations³⁴; Extended Data Figs 4 and 5), which have also been observed beneath modern-day ice streams³⁵, and which typically exhibit convergent flow patterns towards a main ice stream 'trunk'. These landform assemblages are often characterized by abrupt lateral margins that border areas with much shorter subglacial bedforms, or no bedforms at all^{32,33,36,37,42,43} (Extended Data Fig. 5). In some cases, the abrupt margin is marked by features known as ice stream shear margin moraines^{32,33,37,38} (Extended Data Fig. 5). This landform evidence is readily identifiable on satellite imagery and aerial photographs^{32,34,36–39,43} and, in some cases, is further augmented by field investigations that reveal discrete (Boothia-type)³³ erratic dispersal trains or sedimentological evidence from tills that may have been overridden and/or deformed by rapid ice flow. However, there are no ice streams in the inventory that are there only on the basis of sedimentological data and most (>85%) have a clear bedform imprint¹⁴.

At a larger spatial scale, it is known that many ice streams are steered by the underlying topography, often forming marine-terminating outlet glaciers bordered by rock walls^{2–4}. The inventory includes these topographic ice streams, many of which were identified by major cross-shelf troughs and their associated sedimentary depocentres⁴⁰ (Extended Data Fig. 6). Swath bathymetry from within these troughs commonly reveals many of the geomorphological criteria³² described above, such as mega-scale glacial lineations^{34,35,42}.

Given previous work that highlights the importance of 'soft' bedrock geology in influencing ice stream location^{10,18,19}, we also analysed the type of bedrock over which each ice stream was located. We categorized their underlying geology as either (i) predominantly 'soft' sedimentary rocks, (ii) predominantly 'hard' crystalline rocks (intrusive, metamorphic and volcanic rocks) or (iii) those where the spatial footprint of the ice stream extended over a mixture of both soft and hard rocks. This allowed us to calculate the number of different types of ice streams in each broad geological category over time (Extended Data Fig. 3).

Dating Laurentide ice streams. We used the best-available pan-ice sheet margin chronology¹⁵ to bracket the age of the spatial footprints of each ice stream in the inventory¹⁴. The ice margin chronology includes 32 time steps, starting 21.8 kyr ago (18 ¹⁴C (radiocarbon) kyr ago) and ending 5.7 kyr ago (5 ¹⁴C kyr), based on >4,000 dates that are spread across the entire ice sheet bed (Extended Data Fig. 2). The database consists of mainly radiocarbon dates, supplemented with varve and tephra dates, which constrain ice margin positions and shorelines of large glacial lakes. Dates on problematic materials (for example, marl, freshwater shells, lake sediment with low organic carbon content, marine sediment, bulk samples with probable blended ages, and most deposit-feeding molluscs from calcareous substrates) were excluded. Marine-shell dates, a major component, were adjusted for regionally variable marine-reservoir effects on the basis of a large new set of radiocarbon ages on live-collected, pre-bomb molluscs from Pacific, Arctic and Atlantic shores. We use a mixed marine and Northern Hemisphere atmosphere calibration curve, whereas a IntCal98 calibration curve was used in ref. 15.

We used the ice margin chronology to bracket the duration of ice stream activity using methods employed in previous work on individual or small numbers of ice streams^{21,36,41–43} (Extended Data Fig. 7). In some cases, the duration of ice streaming might have been short-lived (only a few hundred years), leaving evidence of a simple 'rubber-stamped' imprint³² of their activity, the spatial extent of which can be readily matched to just one or two ice margin positions (Extended Data Fig. 7a). The more complex landform assemblages of other ice streams (with overprinted mega-scale glacial lineations linked to associated ice marginal features) clearly indicate that they continued to operate during ice margin retreat (Extended Data Fig. 7b); therefore, we fitted the ice stream activity to a series of ice margin positions

over a longer time span (hundreds to thousands of years). Similar patterns are seen for marine-terminating ice streams⁴² (Extended Data Fig. 7c, d).

To account for the inherent uncertainties in the dating (and interpolated ice margin position) and the spatial extent of each ice stream, we provide a maximum possible duration and a minimum duration for each ice stream in the inventory (Fig. 2). In cases in which the interpolated ice margin positions indicated a very short duration of ice streaming, we set the minimum duration to 100 yr. We chose this value because the creation of subglacial bedforms that permit the identification of ice streams probably occurs on timescales of the order of decades³⁵, and attempting to date to a higher precision is meaningless given the dating uncertainties (mainly radiocarbon) and our focus on millennial-scale changes throughout deglaciation.

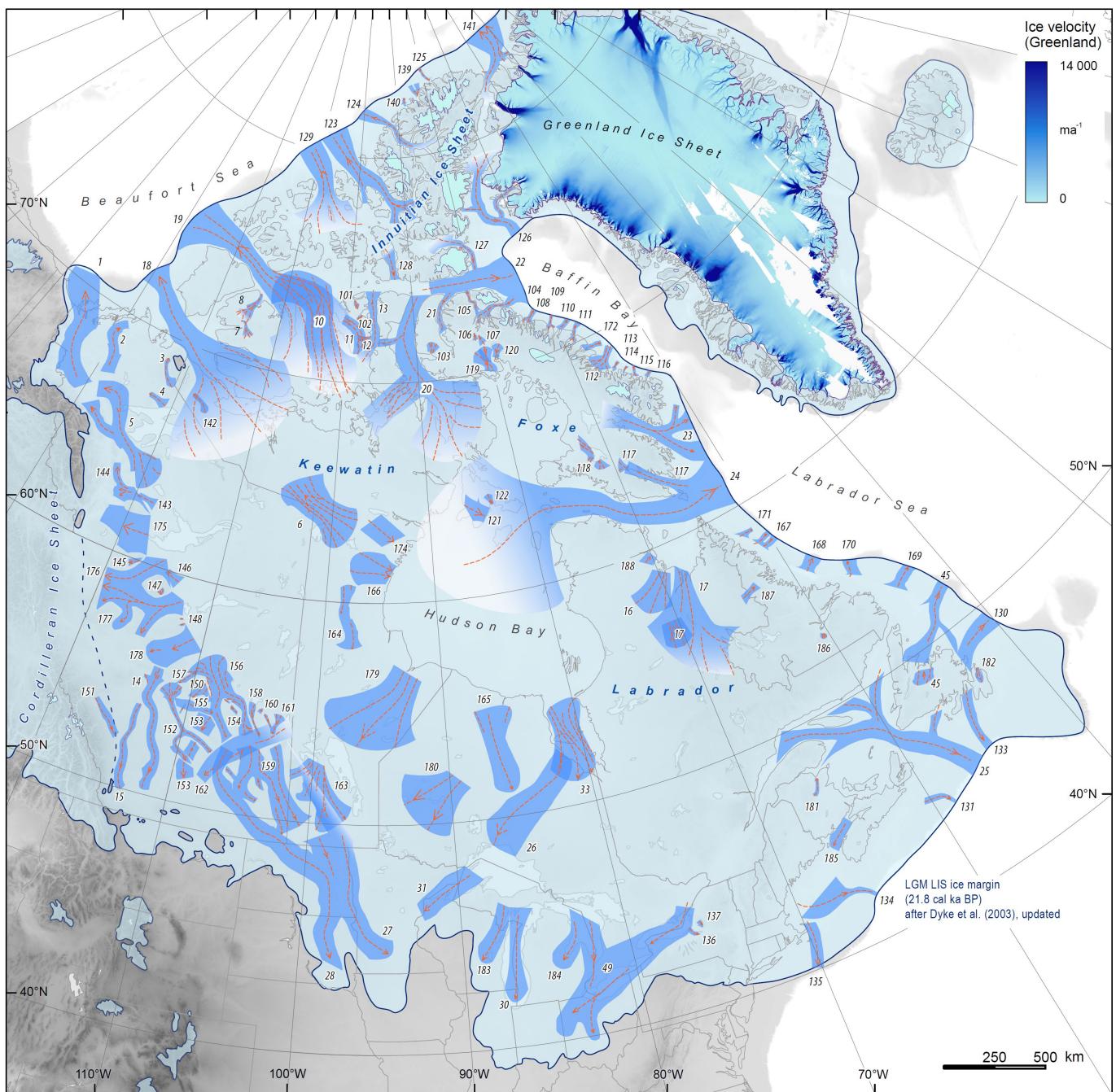
Estimating Laurentide ice stream discharges. Unfortunately, there is no direct way to empirically reconstruct the velocity, and thus discharge, of an ice stream from the evidence it left behind. To provide a simple, first-order estimate of the potential ice discharge from each ice stream for which only the width is known confidently, we used an empirical relationship between the width and the discharge of 81 active ice streams, 50 in Antarctica (Extended Data Fig. 8) and 31 in Greenland (Extended Data Fig. 9). Ice velocities (in units of metres per year) were extracted from recent compilations in Greenland (2008–2009)^{44,45} and Antarctica (2007–2009)^{11,45} and we used these velocity data sets to measure the width (in kilometres) of the ice stream (to the lateral shear margins or exposed rock walls) at the grounding line^{29,46}. Velocity was extracted as a width-averaged value. We then used the highest resolution bed data that was available for Greenland²⁹ and Antarctica³⁰ to calculate the cross-sectional area (in square kilometres) of each ice stream at the grounding line. We calculated the modern ice stream discharge (in cubic kilometres per year) by multiplying the velocity data by the ice-thickness data and integrating the output along the width of the ice stream at the grounding line.

When ice stream data from Antarctica and Greenland are amalgamated (Extended Data Fig. 10), a simple linear regression reveals a weak correlation ($R^2 = 0.39$) between ice stream width and discharge, which we used to predict an order-of-magnitude discharge from the width of each Laurentide ice stream that was active during deglaciation at each dated margin position (Fig. 3d). The regression is clearly influenced by two outliers with extremely high discharge (Pine Island Glacier and Thwaites Glacier, West Antarctica). Without them, the correlation weakens ($R^2 = 0.31$) and our discharge estimates show the same trend, but absolute discharges are lower (see grey shading in Fig. 3d). We considered removing them from the regression, but chose not to and used them in our estimates of Laurentide ice stream discharge (for example, in Figs 3e, f and 4c, d) because they allow us to partly capture some of the more extreme discharges that might be expected in a deglaciating ice sheet. We also extracted the 95% confidence intervals of the regression and used these to estimate a lower and upper range of discharge for an ice stream of given width. However, these confidence intervals under-represent the uncertainty because some assumptions for those confidence intervals (and the general validity of linear regression) are broken: (i) Gaussian noise, (ii) no correlation between residuals of individual data points and (iii) constant variance. Given this, and the clear (and perhaps not surprising) complexity of the relationship between discharge and width for modern-day ice streams (for example, the mean value of linear ice stream flux for Greenlandic ice streams is very different from that for Antarctic ice streams, for which there is a stronger relationship), we also extracted a discharge relationship with a cruder, two-state approximation that avoids the assumptions required for statistically robust application of linear regression. The modern-day ice stream data are from one short time period, and yet we know that ice streams with a fixed width can accelerate (and decelerate) on short (annual–decadal) timescales^{2–4}; however, the extent to which these accelerations and decelerations are sustained over longer (centennial–millennial) timescales is currently unknown. Therefore, we use the simple approach to generate an empirical order-of-magnitude estimate of ice stream discharge from the LIS, averaged over millennial timescales, and note that our empirical results are broadly similar to those generated by numerical modelling^{16,17} (Fig. 3f), albeit typically lower.

To evaluate our empirical estimates of ice stream discharge in relation to ice sheet volume (see, for example, Figs 3d, e and 4), we extracted ice sheet volume from the mean of an ensemble of best-performing model runs from a previously published data-calibrated numerical model¹⁶. Uncertainties associated with the modelled ice volumes (see ref. 16) are an order of magnitude less than those associated with our estimates of ice stream discharge and are not shown (for example, in Fig. 4). We use this same model to compare our empirical estimates of ice stream discharge against those generated in a numerical model of the LIS, with streaming discharge extracted from an ensemble of best-performing model runs at 100-yr time steps during deglaciation from 21.8 kyr ago to 5.7 kyr ago and ensemble standard deviation in ice stream discharge shown as shading around the mean (Fig. 3f). The weighted ensemble mean from this model shows

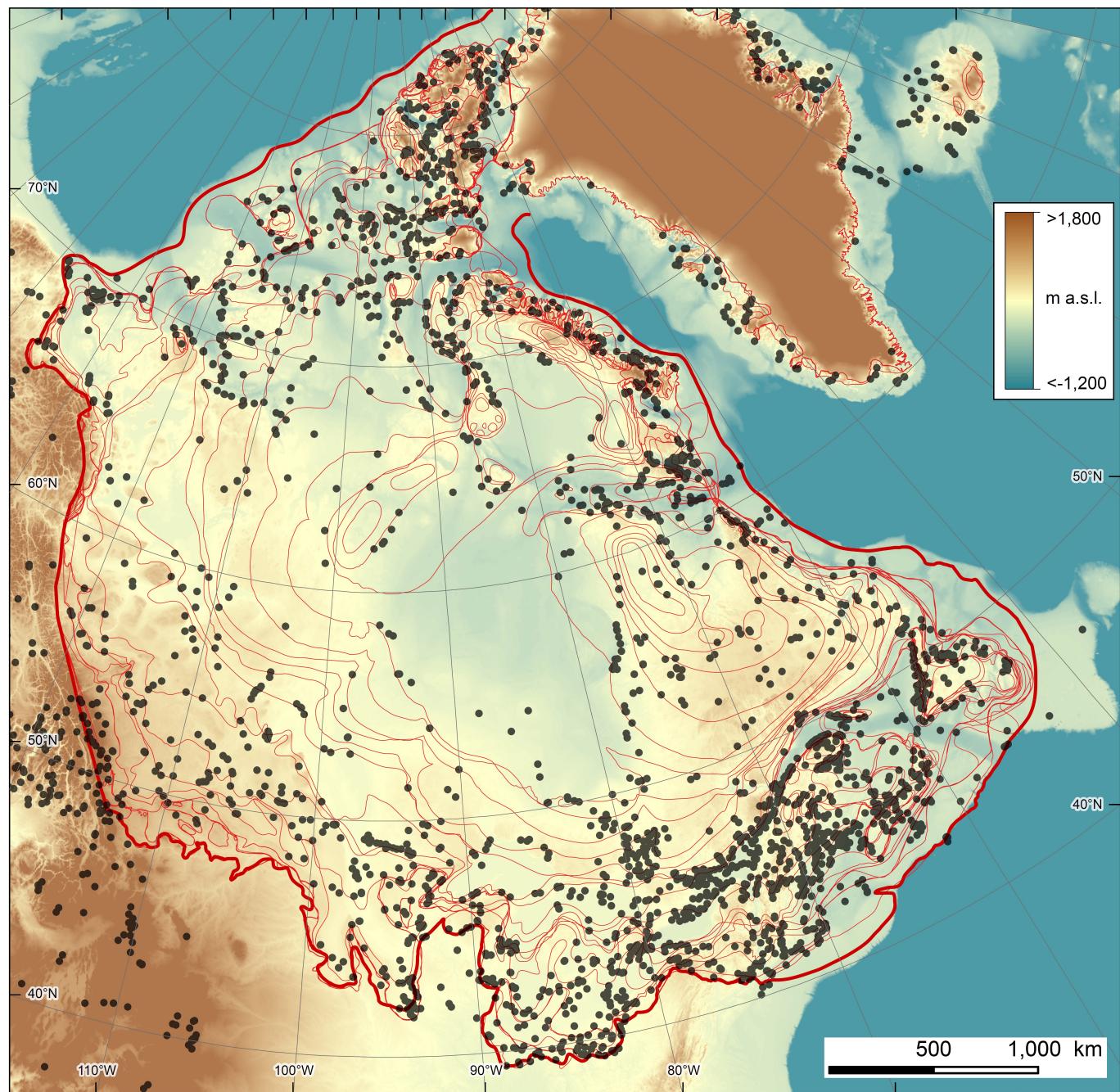
a similar trend of decreasing discharge from ice streams (Fig. 3f), but with higher discharges and greater variability. This is to be expected because our estimates based on modern-day ice stream discharges may not capture the full range of ice stream behaviour during deglaciation of a mid-latitude ice sheet (for example, we have no modern analogue of an extensive land-terminating margin overlying soft sediments). The numerical modelling also imposes a data-calibrated reduction in ice stream discharge around the Hudson Strait region just before H1 to facilitate a dynamic destabilization during this event. This imposed reduction is reflected in the reduced ice stream discharge in that model for a few thousand years before 17 kyr ago and a temporary increase thereafter (Figs 3f and 4c, d). It is also reflected in the low modelled streaming fraction at volumes of about $2.4 \times 10^7 \text{ km}^3$ (see Fig. 4d).

31. Patterson, W. S. B. *The Physics of Glaciers* 3rd edn, 301 (Pergamon, 1994).
32. Stokes, C. R. & Clark, C. D. Geomorphological criteria for identifying Pleistocene ice streams. *Ann. Glaciol.* **28**, 67–74 (1999).
33. Dyke, A. S. & Morris, T. F. Drumlin fields, dispersal trains, and ice streams in Arctic Canada. *Can. Geogr.* **32**, 86–90 (1988).
34. Clark, C. D. Mega-scale glacial lineations and cross-cutting ice flow landforms. *Earth Surf. Process. Landf.* **18**, 1–29 (1993).
35. King, E. C., Hindmarsh, R. C. A. & Stokes, C. R. Formation of mega-scale glacial lineations observed beneath a West Antarctic ice stream. *Nature Geosci.* **2**, 585–588 (2009).
36. Clark, C. D. & Stokes, C. R. Extent and basal characteristics of the M'Clintock Channel Ice Stream. *Quat. Int.* **86**, 81–101 (2001).
37. Hodgson, D. A. Episodic ice streams and ice shelves during retreat of the northwesternmost sector of the late Wisconsinan Laurentide Ice Sheet over the central Canadian Arctic Archipelago. *Boreas* **23**, 14–28 (1994).
38. Stokes, C. R. & Clark, C. D. Ice stream shear margin moraines. *Earth Surf. Process. Landf.* **27**, 547–558 (2002).
39. Stokes, C. R. Identification and mapping of palaeo-ice stream geomorphology from satellite imagery: implications for ice stream functioning and ice sheet dynamics. *Int. J. Remote Sens.* **23**, 1557–1563 (2002).
40. Batchelor, C. L. & Dowdeswell, J. A. The physiography of high Arctic cross-shelf troughs. *Quat. Sci. Rev.* **92**, 68–96 (2014).
41. Stokes, C. R. & Clark, C. D. Palaeo-ice streams. *Quat. Sci. Rev.* **20**, 1437–1457 (2001).
42. O’Cofaigh, C., Dowdeswell, J. A., Evans, J. & Larter, R. D. Geological constraints on Antarctic palaeo-ice stream retreat. *Earth Surf. Process. Landf.* **33**, 513–525 (2008).
43. Stokes, C. R., Clark, C. D. & Storrar, R. D. Major changes in ice stream dynamics during deglaciation of the north-western margin of the Laurentide Ice Sheet. *Quat. Sci. Rev.* **28**, 721–738 (2009).
44. Joughin, I., Smith, B., Howat, I. & Scambos, T. MEaSUREs Greenland Ice Sheet Velocity Map from InSAR Data. <https://nsidc.org/data/nsidc-0478> (National Snow and Ice Data Center, 2010).
45. Joughin, I., Smith, B., Howat, I., Scambos, T. & Moon, T. Greenland flow variability from ice-sheet-wide velocity mapping. *J. Glaciol.* **56**, 415–430 (2010).
46. Rignot, E., Mouginot, J. & Scheuchl, B. MEaSUREs InSAR-Based Antarctica Ice Velocity Map. <http://nsidc.org/data/NSIDC-0484> (National Snow and Ice Data Center, 2011).
47. Global Digital Evaluation Model (GTOPO30). US Geological Survey, EROS Data Center Distributed Active Archive Center (EDC DAAC) <http://eros.usgs.gov/> (2004).
48. Briner, J. P., Miller, G. H., Davis, P. T. & Finkel, R. C. Cosmogenic radionuclides from fjord landscapes support differential erosion by overriding ice sheets. *Geol. Soc. Am. Bull.* **118**, 406–420 (2006).
49. Shaw, J. et al. A conceptual model of the deglaciation of Atlantic Canada. *Quat. Sci. Rev.* **19**, 959–980 (2000).
50. Kleman, J. et al. North American Ice Sheet build-up during the last glacial cycle, 115–21 kyr. *Quat. Sci. Rev.* **29**, 2036–2051 (2010).
51. Lakeman, T. R. & England, J. H. Palaeo-glaciological insights from the age and morphology of the Jesse moraine belt, western Canadian Arctic. *Quat. Sci. Rev.* **47**, 82–100 (2012).
52. Lakeman, T. R. & England, J. H. Late Wisconsinan glaciation and postglacial relative sea-level change on western Banks Island, Canadian Arctic Archipelago. *Quat. Res.* **80**, 99–112 (2013).
53. Jakobsson, M. et al. Arctic Ocean glacial history. *Quat. Sci. Rev.* **92**, 40–67 (2014).
54. Nixon, F. C. & England, J. H. Expanded Late Wisconsinan ice cap and ice sheet margins in the western Queen Elizabeth Islands, Arctic Canada. *Quat. Sci. Rev.* **91**, 146–164 (2014).
55. Jakobsson, M. et al. The International Bathymetric Chart of the Arctic Ocean (IBCAO) Version 3.0. *Geophys. Res. Lett.* **39**, L12609 (2012).
56. ArcticNet. <http://www.omg.unb.ca/Projects/Arctic/> [accessed 1 December 2013] (2013).
57. Margold, M., Stokes, C. R. & Clark, C. D. Ice streams in the Laurentide Ice Sheet: identification, characteristics and comparison to modern ice sheets. *Earth Sci. Rev.* **143**, 117–146 (2015).
58. Rignot, E., Mouginot, J. & Scheuchl, B. Antarctic grounding line mapping from differential satellite radar interferometry. *Geophys. Res. Lett.* **38**, L10504 (2011).



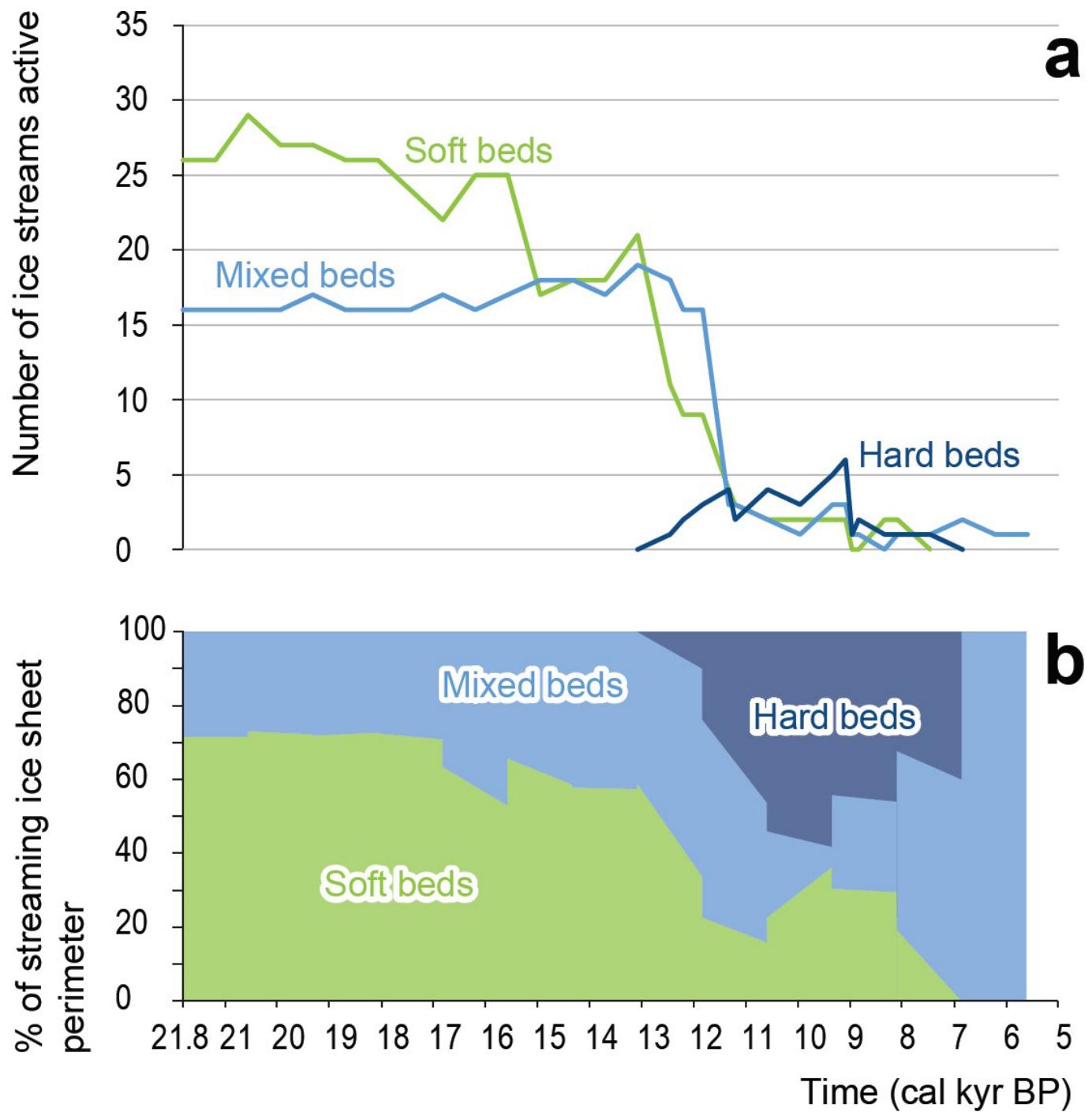
Extended Data Figure 1 | Location of 117 ice streams from a recently compiled inventory based on previous work and systematic mapping across the LIS bed. Ice streams are shaded dark blue and numbered according to their inventory¹⁴ number. Orange dashed lines with arrows

show ice stream flow direction. Modern-day ice velocity is shown for Greenland^{44,45}. Underlying topography from GTOPO30 digital elevation data⁴⁷.



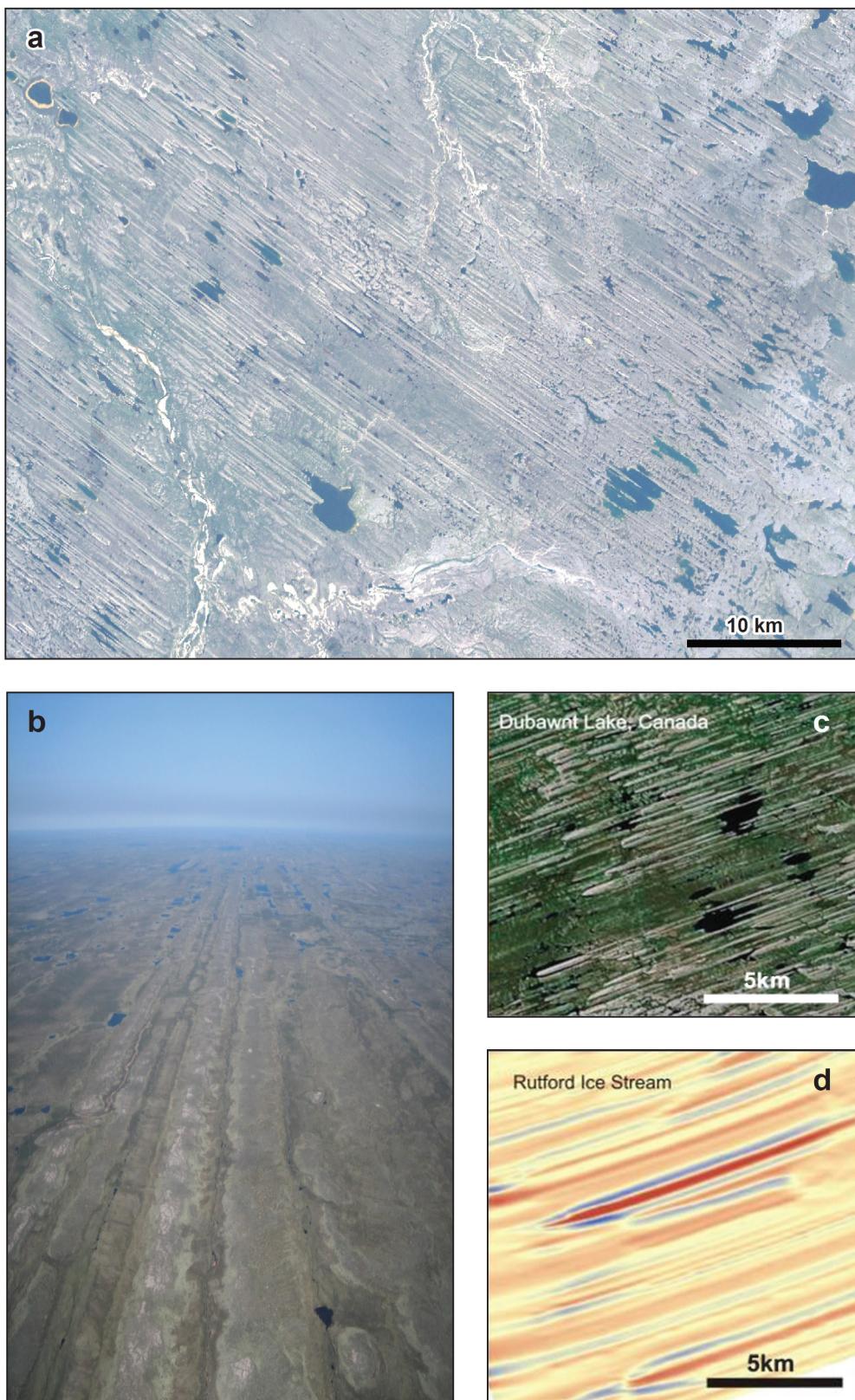
Extended Data Figure 2 | Distribution of dates and interpolated ice margin positions. These ice margin positions¹⁵ (thin red lines) are based on dates (black dots) that we used to bracket the age of the spatial footprint of each Laurentide ice stream (Extended Data Fig. 7). The thick

red line shows the updated LGM ice margin (following recent work^{48–54}). Underlying topography from GTOPO30 digital elevation data⁴⁷. m a.s.l., metres above sea level.



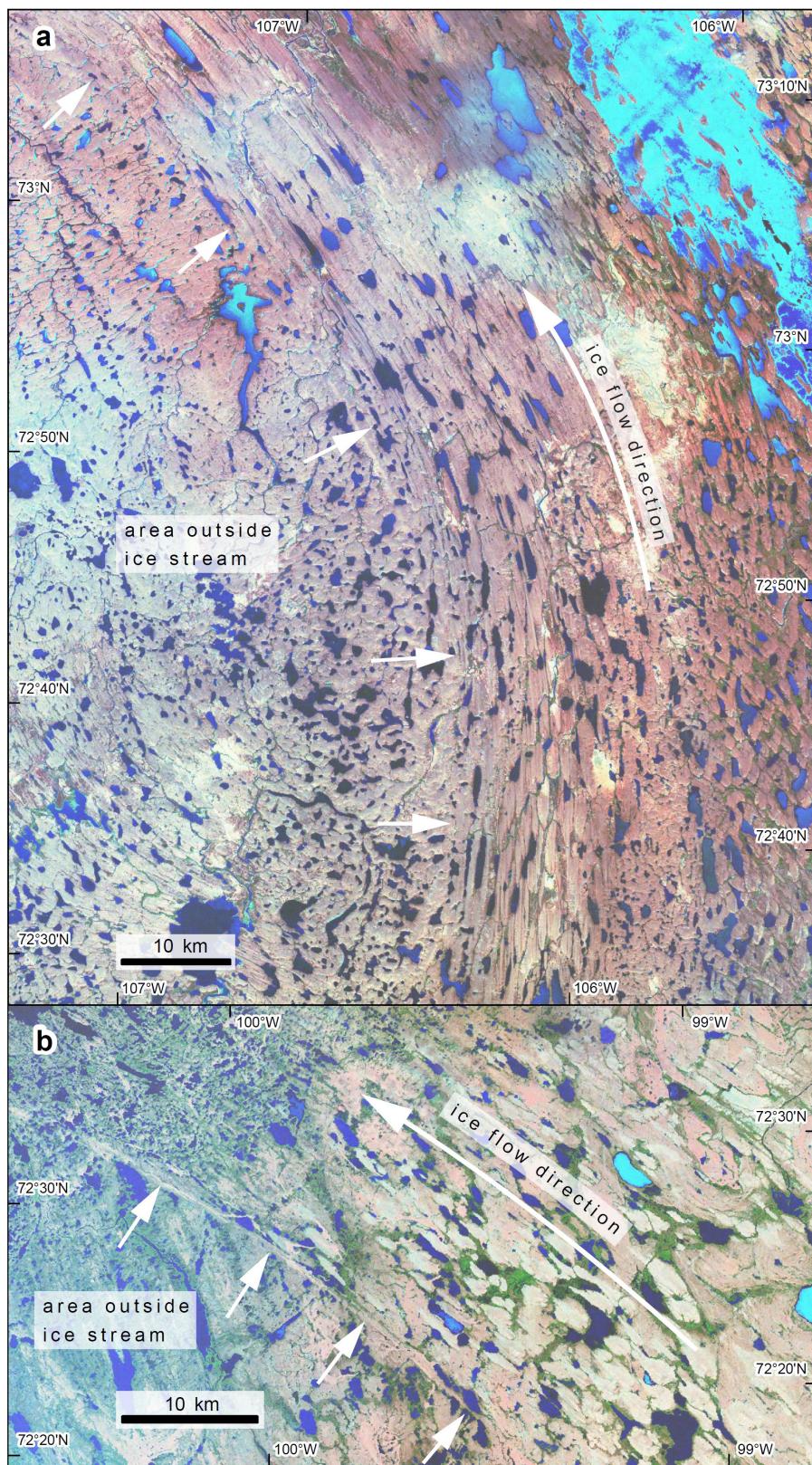
Extended Data Figure 3 | Number of Laurentide ice streams and the percentage of the margin they drained over time, classified according to their underlying geology. **a**, A rapid decrease in the number of ice streams is observed after about 12 kyr ago (Fig. 3a), which is linked with the retreat of the LIS onto the hard crystalline rocks of the Canadian Shield¹⁹

(cal kyr BP refers to thousands of calendar years before present). **b**, However, several large, wide ice streams were active over the hard bed geology (see, for example, refs 10, 21), and they drained a large percentage of the perimeter of the ice sheet.



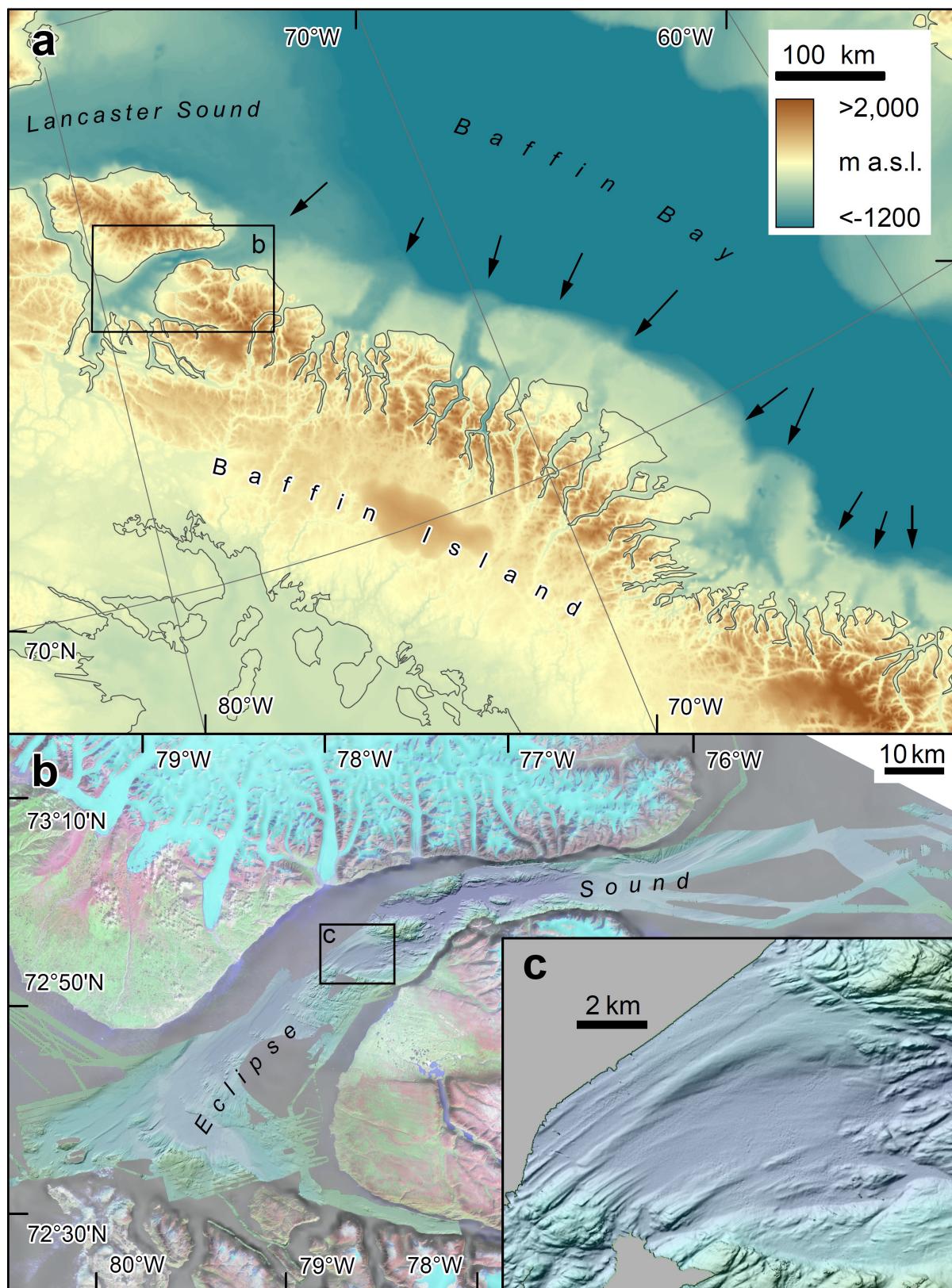
Extended Data Figure 4 | Megascale glacial lineations³⁴ on the bed of the Dubawnt Lake ice stream²¹, central Canada. **a–c**, These features are a characteristic geomorphological signature of ice streaming and are readily identifiable on Landsat satellite imagery (a, c) and oblique aerial photography (b) of the ice stream bed (number 6 in Extended Data Fig. 1).

d, Similar features have been detected beneath Rutherford ice stream, West Antarctica³⁵. Landsat imagery courtesy of the US Geological Survey Earth Resources Observation Science Centre; photograph by C.R.S. Images in c and d modified from ref. 35.



Extended Data Figure 5 | Landsat imagery of lateral shear margin moraines in the Canadian Arctic Archipelago. **a, b**, Beds of the McClintock Channel ice stream^{36,37}, Victoria Island (**a**; number 10 in Extended Data Fig. 1) and the Crooked Lake Ice Stream³³, Prince of Wales

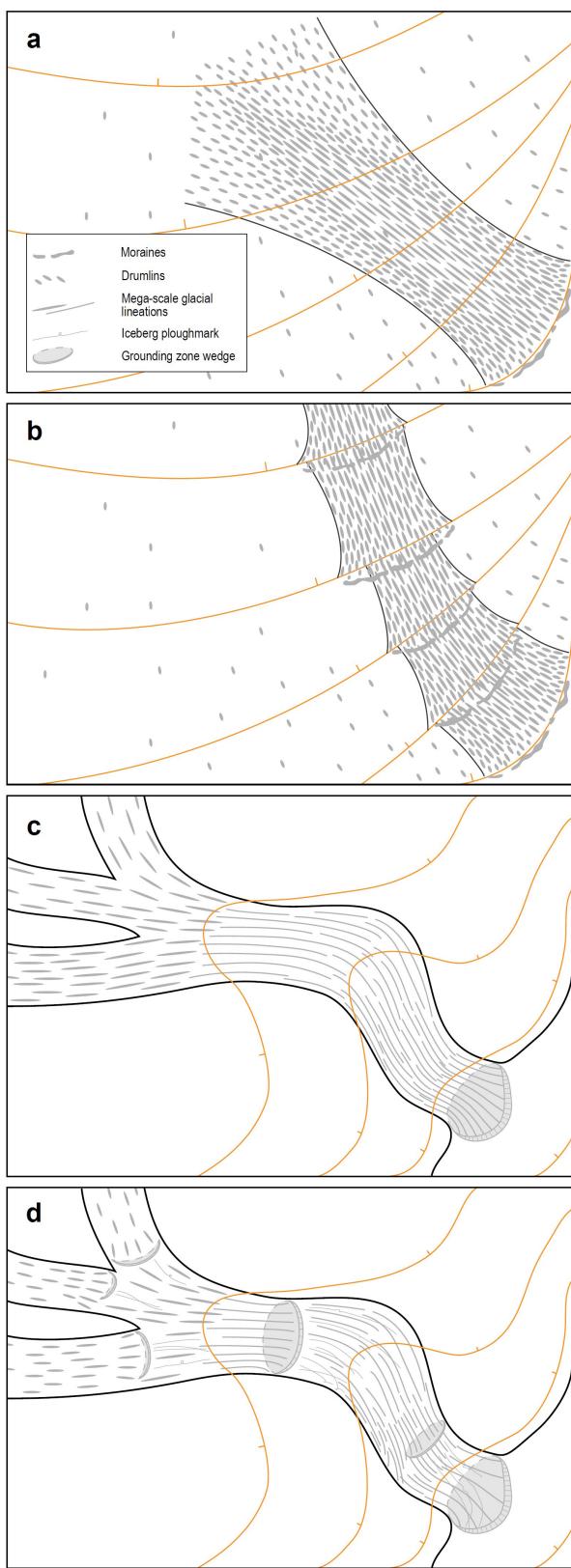
Island (**b**; number 11 in Extended Data Fig. 1). Note the abrupt lateral margins (marked by white arrows) of the assemblage of mega-scale glacial lineations that are, in places, marked by lateral shear margin moraines³⁸.



Extended Data Figure 6 | Bathymetric data showing cross-shelf troughs and a well-preserved bedform imprint from a submarine setting.

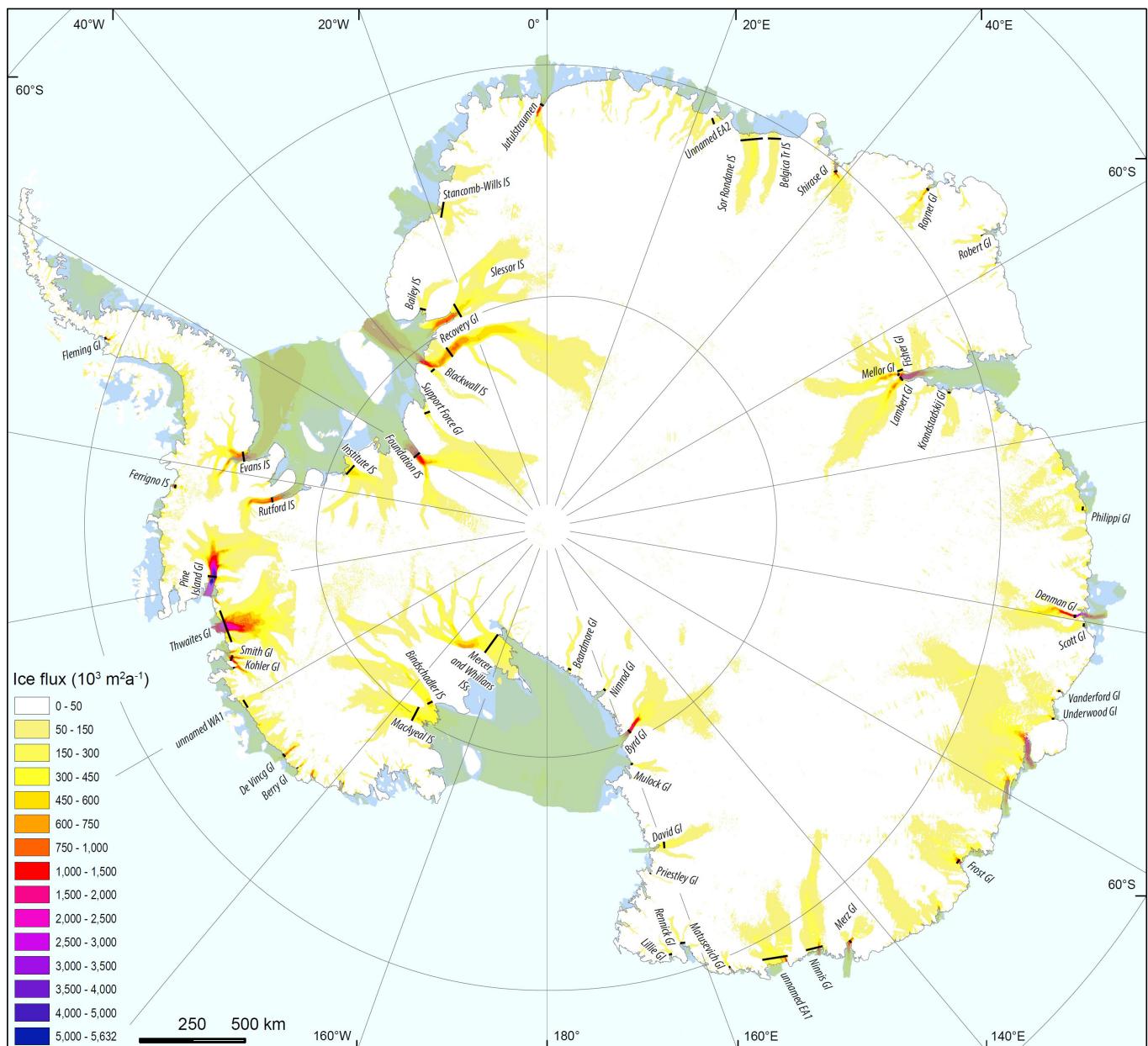
a, Cross-shelf troughs formed by ice streams fed by convergence of ice flow from several fjords along the east coast of Baffin Island. **b, c**, Drumlins and mega-scale glacial lineations on the floor of Eclipse Sound (b; boxed

area in a) and a close-up view of one particular region where subglacial bedforms are well developed (c; boxed area in b). High-resolution swath bathymetry data are from IBCAO⁵⁵ (a), and from IBCAO⁵⁵ and ArcticNet⁵⁶ (b, c). Figure redrawn from ref. 57.



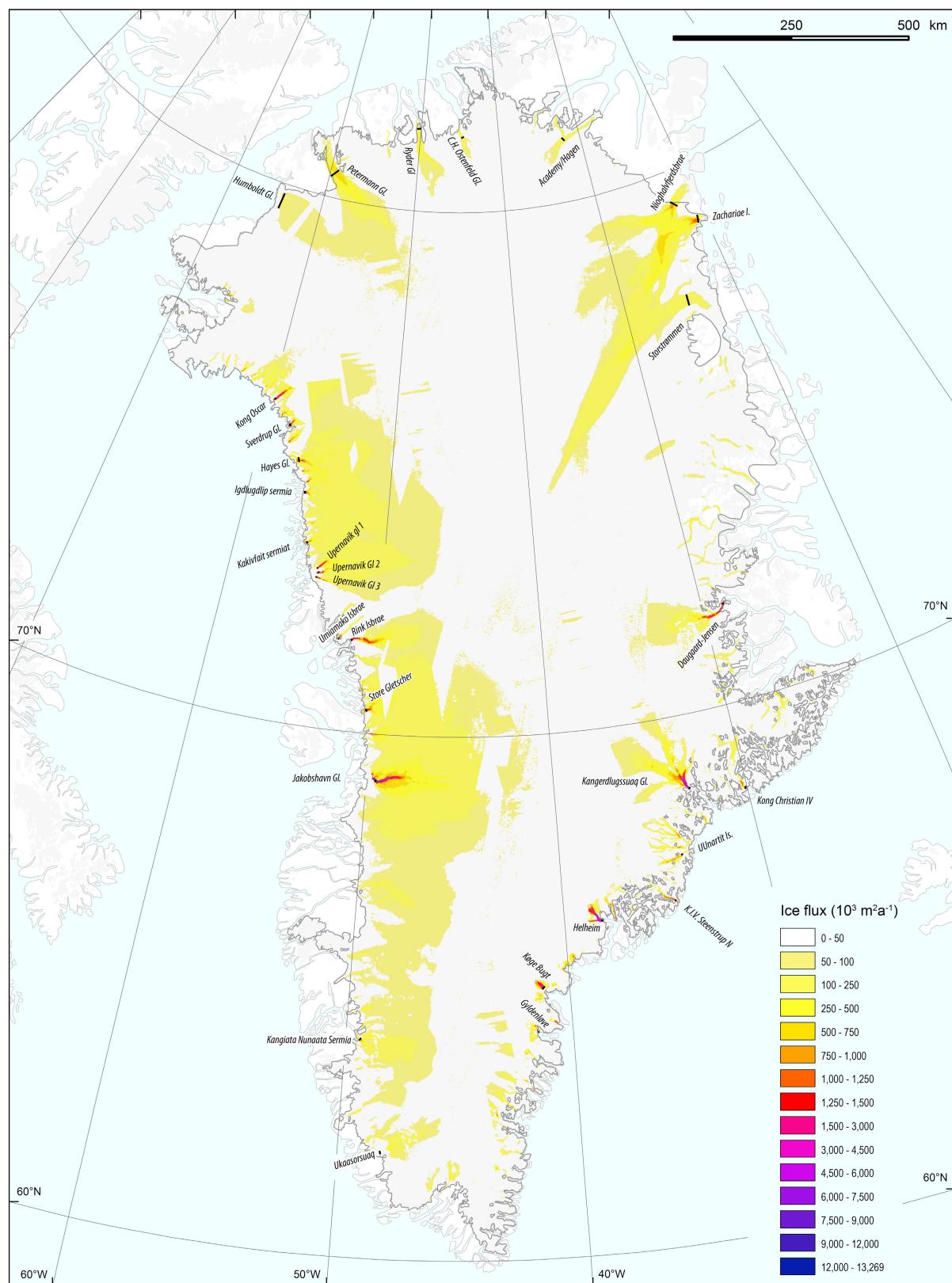
Extended Data Figure 7 | See next page for figure caption.

Extended Data Figure 7 | Schematic demonstrating the method used to bracket the age of the spatial footprint of Laurentide ice streams in both terrestrial and marine settings. These methods have been used extensively in previous work, but usually on small samples of ice streams (see, for example, refs 20, 21, 36, 37, 42, 43). **a**, In some cases, terrestrial ice streams are active, but then deactivate (switch off) as the ice margin retreats, which enables them to be bracketed between a small number of dated ice margins and implies a short duration of operation. **b**, In other cases, ice streams remain active during deglaciation and continually remould their landform assemblage, leaving a more complicated, time-integrated landform record, often with a series of overprinted landforms, which implies a longer duration of operation. **c, d**, The same scenarios as in **a** and **b**, respectively, but for a topographically controlled marine-terminating ice stream.



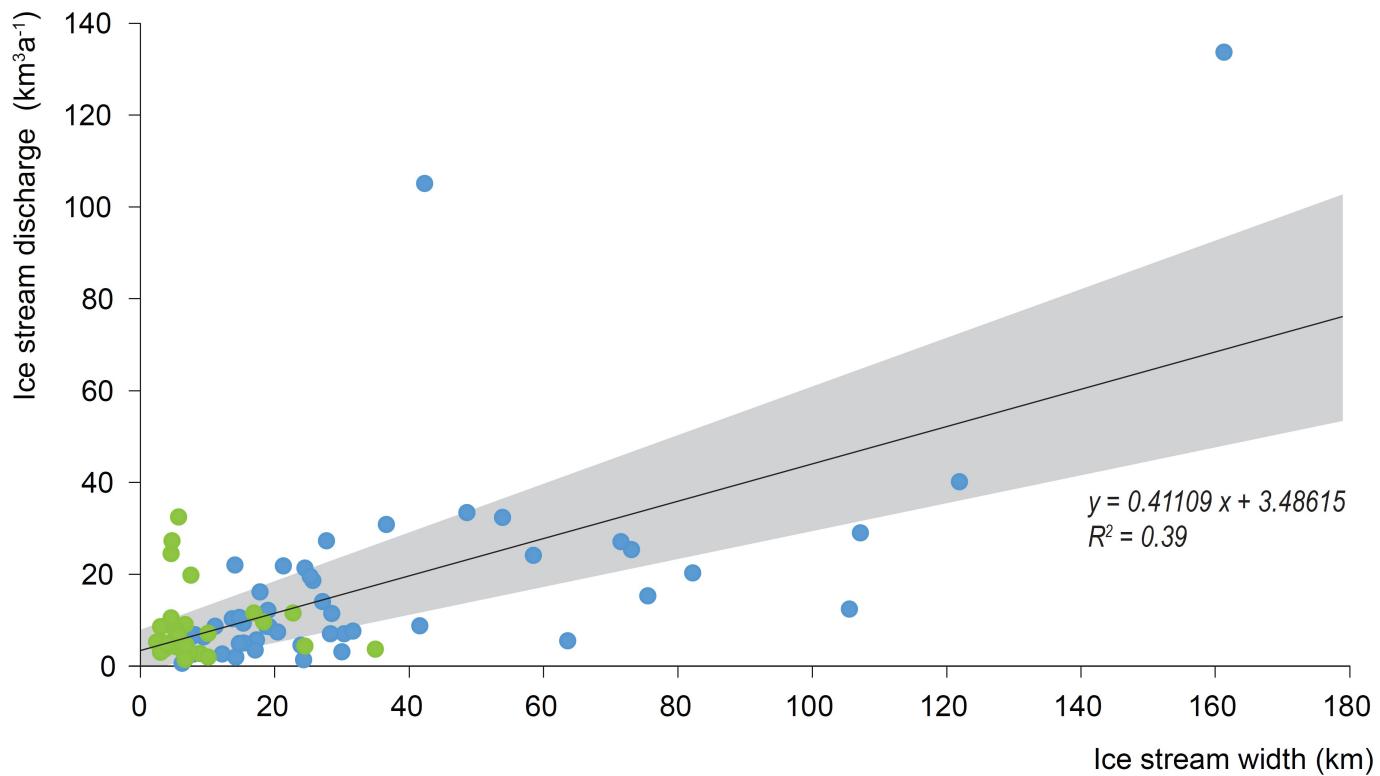
Extended Data Figure 8 | Location of ice streams in Antarctica where discharge was estimated. Discharge was estimated from existing data sets of velocity^{11,46}, grounding-line position⁵⁸ and ice thickness³⁰. Regression analysis reveals a weak relationship between the width and discharge of

the ice streams (Extended Data Fig. 10), which we used to estimate the discharge of Laurentide ice streams for which we know only their width (see Methods).



Extended Data Figure 9 | Location of ice streams in Greenland where discharge was estimated. Discharge was estimated from existing data sets of velocity^{44,45}, grounding-line position^{44,45} and ice thickness²⁹. Regression analysis reveals a weak relationship between the width and

discharge of the ice streams (Extended Data Fig. 10), which we used to estimate the discharge of Laurentide ice streams for which we know only their width (see Methods).



Extended Data Figure 10 | Relationship between ice stream discharge and width for 81 active ice streams in Antarctica and Greenland.
Discharge calculations derived from velocity data from 2008–2009

(Greenland, green dots)^{44,45} and 2007–2009 (Antarctica, blue dots)^{11,46,58}.
Grey shading shows the 95% confidence interval of the linear regression.
Measured ice stream locations are shown in Extended Data Figs 8 and 9.

Moralistic gods, supernatural punishment and the expansion of human sociality

Benjamin Grant Purzycki¹, Coren Apicella², Quentin D. Atkinson^{3,4}, Emma Cohen^{5,6}, Rita Anne McNamara⁷, Aiyana K. Willard⁸, Dimitris Xygalatas^{9,10,11}, Ara Norenzayan⁷ & Joseph Henrich^{7,12,13}

Since the origins of agriculture, the scale of human cooperation and societal complexity has dramatically expanded^{1,2}. This fact challenges standard evolutionary explanations of prosociality because well-studied mechanisms of cooperation based on genetic relatedness, reciprocity and partner choice falter as people increasingly engage in fleeting transactions with genetically unrelated strangers in large anonymous groups. To explain this rapid expansion of prosociality, researchers have proposed several mechanisms^{3,4}. Here we focus on one key hypothesis: cognitive representations of gods as increasingly knowledgeable and punitive, and who sanction violators of interpersonal social norms, foster and sustain the expansion of cooperation, trust and fairness towards co-religionist strangers^{5–8}. We tested this hypothesis using extensive ethnographic interviews and two behavioural games designed to measure impartial rule-following among people ($n = 591$, observations = 35,400) from eight diverse communities from around the world: (1) inland Tanna, Vanuatu; (2) coastal Tanna, Vanuatu; (3) Yasawa, Fiji; (4) Lovu, Fiji; (5) Pesqueiro, Brazil; (6) Pointe aux Piments, Mauritius; (7) the Tyva Republic (Siberia), Russia; and (8) Hadzaland, Tanzania. Participants reported adherence to a wide array of world religious traditions including Christianity, Hinduism and Buddhism, as well as notably diverse local traditions, including animism and ancestor worship. Holding a range of relevant variables constant, the higher participants rated their moralistic gods as punitive and knowledgeable about human thoughts and actions, the more coins they allocated to geographically distant co-religionist strangers relative to both themselves and local co-religionists. Our results support the hypothesis that beliefs in moralistic, punitive and knowing gods increase impartial behaviour towards distant co-religionists, and therefore can contribute to the expansion of prosociality.

Among the other factors^{2–4,7} that influence the emergence of human ultrasociality and complex societies, the diffusion of explicit beliefs in increasingly moralistic, punitive and knowledgeable gods may have played a crucial role^{6,7}. People may trust in, cooperate with and interact fairly within wider social circles, partly because they believe that knowing gods will punish them if they do not. Additionally, through increased frequency and consistency in belief and behaviour sets, commitments to the same gods coordinate people's expectations about social interactions^{5–9}. Moreover, the social radius within which people are willing to engage in behaviours that benefit others at a cost to themselves may enlarge as gods' powers to monitor and punish increase¹⁰. To account for the emergence of these patterns, some evolutionary approaches to religion have theorized that cultural evolution may have

harnessed and exploited aspects of our evolved psychology, such as mentalizing abilities, dualistic tendencies and sensitivity to norm compliance, to gradually assemble configurations of supernatural beliefs that promote greater cooperation and trust within expanding groups, leading to greater success in intergroup competition. Of course, given that cultural evolution can produce self-reinforcing stable patterns of beliefs and practices, these supernatural agent concepts may also have been individually favoured within groups due to mechanisms related to signalling, reputation and punishment^{5–9,11,12}. Over time, these deities spread culturally and came to dominate the modern world religions like Christianity, Islam and Hinduism. Such traditions eventually came to account for a large proportion of the world's population^{6,7,13,14} (see Supplementary Information section S1). Here we directly test one specific hypothesis: conceptions of moralistic and punitive gods that know people's thoughts and behaviours promote impartiality towards distant co-religionists, and as a result contribute to the expansion of sociality.

At the societal level, several lines of converging evidence are consistent with this hypothesis. For example, after controlling for key correlates, analyses of cross-cultural data sets show that larger and more politically complex societies tend to have more supernatural punishment and moralistic deities^{5,15}, and historical analyses in one geographic region show that precursors to supernatural punishment beliefs precede social complexity¹⁶. However, this data derives from qualitative ethnographies of entire societies; a more focused, direct and systematic cross-cultural assessment of what individuals think their gods care about, and whether or not people explicitly or implicitly view their gods as concerned with norms of interpersonal social behaviour (termed here as 'morality'^{17,18}; see Supplementary Information section S4.2) has only recently begun^{18–20}. Analyses of cross-national databases (for example, the World Values Survey) reveal positive relationships between beliefs in hell, beliefs in gods' power to punish, and various self-reported prosocial behaviours^{21,22}. Although valuable, these lines of research primarily rely on survey questions not specifically designed to address the research question we are interested in. Moreover, they rely on samples drawn broadly from nation states, thus excluding small-scale societies that are crucial for assessing questions about the expansion of prosociality.

At the individual level, two types of behavioural studies are also consistent with this hypothesis, but each has crucial limitations. First, laboratory experiments show that exposure to religious reminders increases generosity and decreases cheating among religious believers^{23–25}. However, as is the case for most psychological experiments, the vast majority of these studies rely on Western, Christian-majority samples, limiting their generalizability²⁶. Second, in one

¹Centre for Human Evolution, Cognition, and Culture, University of British Columbia, 1871 West Mall, Vancouver, British Columbia V6T 1Z2, Canada. ²Department of Psychology, University of Pennsylvania, Solomon Laboratories, 3720 Walnut Street, Philadelphia, Pennsylvania 19104-6241, USA. ³Department of Psychology, University of Auckland, Human Sciences Building, 10 Symonds Street, Auckland 1010, New Zealand. ⁴Max Planck Institute for the Science of Human History, Kahlaische Strasse 10, D-07745 Jena, Germany. ⁵Institute of Cognitive and Evolutionary Anthropology, University of Oxford, 64 Banbury Road, Oxford OX2 6PN, UK. ⁶Wadham College, University of Oxford, Parks Road, Oxford, OX1 3PN, UK. ⁷Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, British Columbia V6T 1Z4, Canada. ⁸Culture, and Development Laboratory, Department of Psychology, The University of Texas at Austin, 1 University Station #A8000, Austin, Texas 78712-0187, USA. ⁹Department of Anthropology, University of Connecticut, 354 Mansfield Road, Unit 1176, Storrs, Connecticut 06029, USA. ¹⁰Interacting Minds Centre, Aarhus University, Jens Chr. Skous Vej 4, building 1483, DK-8000, Aarhus, Denmark. ¹¹LEVYNA, Masaryk University, Brno 60200, Czech Republic. ¹²Department of Economics, University of British Columbia, 2136 West Mall, Vancouver, British Columbia V6T 1Z4, Canada. ¹³Department of Human Evolutionary Biology, Harvard University, 11 Divinity Ave, Cambridge, Massachusetts 02138, USA.

Table 1 | Site descriptive statistics

Site	Researcher	Economy	Moralistic god	Local god or spirit	n	Females	Age	Material insecurity
Coastal Tanna [§]	Atkinson	Horticulture	Christian god	Garden spirit (<i>Tupunus</i>)	44	23	35.02 (14.13)	0.22 (0.36)
Hadza	Apicella	Hunting	Celestial figure (<i>Haine</i>) [#]	Sun (<i>Ishoko</i>) [#]	68	31	39.82 (12.08)	0.82 [‡] (0.36)
Inland Tanna [§]	Atkinson	Horticulture	<i>Kalpapan</i> (traditional)	Garden spirit (<i>Tupunus</i>)	76	38	37.00 (16.17)	0.26 (0.38)
Lovu	Willard	Wage labour	Hindu <i>Bhagwan</i>	None available	76	52	44.56 (16.94)	0.83 (0.33)
Mauritius	Xygalatas	Wage labour and farming	Hindu <i>Shiva</i>	Spirit/soul/ghost (<i>Nam</i>)	94	27	36.56 (15.05)	0.39 (0.35)
Pesqueiro	Cohen	Wage labour	Christian god	Virgin Mary	77	40	34.12 (13.08)	0.86 (0.24)
Tyva Republic	Purzycki	Wage labour and herding	Buddha <i>Burgan</i>	Spirit-masters (<i>Cher eezi</i>)	81	58	33.53 (12.52)	0.47 (0.28)
Yasawa	McNamara	Fishing and farming	Christian god	Ancestor spirits (<i>Kalou-vu</i>)	75	41	38.04 (15.91)	0.50 (0.40)
				Grand mean	73.88	—	37.34	0.55

Means indicated (standard deviations in parentheses). See Extended Data Fig. 1 for a map of field sites.

[§]One individual was removed from the local co-religionist game due to coin visibility.

[#]These two gods closely overlap in conception.

[‡]Answer options were “yes”, “no” or “I don’t know”.

field study²⁷ across 15 diverse societies of foragers, pastoralists and horticulturalists, adherence to Christianity or Islam predicted greater fairness in economic games relative to adherence to local/traditional religions. This study, however, lacked precise measures for our theoretically important components of beliefs about gods’ minds—punishment, knowledge and moralism. Moreover, these studies did not consider the religious affiliation of the anonymous recipients of players’ monetary decisions. It is therefore unclear whether these findings explain the expansion of prosociality specifically towards geographically distant co-religionists.

Addressing these limitations, we combined two behavioural experiments with detailed ethnographic interviews to assess whether participants who report that their moralistic gods are punishing and more knowledgeable about human thought and behaviour are more likely to impartially allocate money to anonymous, geographically distant co-religionists over both themselves and their local community^{6,7}. In five of the sites, we also tested whether religious priming associated with moralistic gods had effects on gameplay, but these had no overall effect (see Supplementary Information sections S2.2.2 and S6.2).

We tested these predictions with a sample of 591 participants (310 females; observations = 35,400; Table 1 and Extended Data Fig. 1) from eight diverse communities, including hunter-gatherers, horticulturalists, herders and farmers, as well as fully market-integrated populations engaged in wage labour or operating small businesses. The participants adhere to a variety of world religious traditions including Christianity, Hinduism and Buddhism, and report beliefs in an immense range of local supernatural agents, including spirit-masters, saints, ancestors, animistic beings, anthropomorphic celestial deities, garden spirits, and ghosts (Supplementary Information section S3).

To measure favouritism towards oneself and local community under maximally anonymous conditions, we modified the random allocation game^{9,28,29}. In this game (Fig. 1), participants play in private with 30 coins, two cups and a fair die with three sides of one colour and three sides of another colour. In the experiment, the participant’s job is to allocate each coin to one of the two cups. First, they mentally choose one of the cups and then roll the die. If one coloured side comes up, players are instructed to put the coin into the cup they mentally

chose. If the die comes up the other colour, people are instructed to put the coin into the opposite cup from the one they chose. Of course, as cup selection occurs only mentally, participants can overrule the die in favour of one of the cups without anyone else observing their decision. If people play by the rules and thereby allocate the coins impartially, the mean number of coins in each cup should be 15, and the distribution around this average will be binomial. This allows us to test for systematic deviations from this distribution (Supplementary Information section S2.2).

Participants played two counterbalanced games for a total of 60 coin allocations per person (Fig. 1). In one game, the local co-religionist game, participants chose between a cup assigned to an unspecified anonymous co-religionist from their local community and a cup assigned to an anonymous co-religionist living in a geographically distant community that does not regularly interact with the player’s community. In the other game, the self game, participants chose between a cup for themselves and a cup for another anonymous distant co-religionist. In order to control for any effects of ethnicity³⁰ and nationality, both local and distant co-religionists were of the same ethnic group and nationality as the participant.

Participants understood that money put into the cups would be given to the people they represented, including themselves, and we actually distributed allocations to participants and randomly selected people described by the cups (that is, there was no deception). After gameplay, we asked each participant a battery of questions, including a series of counterbalanced questions about two locally relevant deities (Supplementary Information section S2).

To assess the gods’ relative moral concern, we conducted preliminary ethnographic interviews in each site to identify the most moralistic deities (that is, ‘moralistic gods’), as well as locally salient, relatively less moralistic, ‘local gods’ or spirits. We verified the degree to which gods care about morality with a free-list task asking about gods’ concerns¹⁹ and scales created to measure how important participants claim punishment, theft, murder and deceit are to these supernatural beings. We measured gods’ punishment and knowledge, using the mean of two, two-item, easy-to-understand scales with dichotomous responses. The target gods associated with games were rated significantly more moralistic, knowledgeable and punitive than local gods (see Extended Data Figs 2 and 3; Supplementary Information section S4). We also aggregated gods’ punishment and knowledge scores by averaging all four dichotomous responses, labelled ‘punishment–knowledge combined’ in Table 2. These measures are our key theoretical predictors for game allocations.

Figure 2 displays the effect of punishment for moralizing gods, without any controls, and reveals the impact of “I don’t know” answers which were otherwise excluded from our analyses below. When people report not knowing if a god punishes, they put considerably fewer coins in the cups for distant co-religionists in both games (local co-religionist game: $M = 12.97$, s.d. = 4.33; self game: $M = 12.50$, s.d. = 4.15) than those who consistently report that their god punishes (local co-religionist game: $M = 14.58$, s.d. = 3.24; self game: $M = 14.53$,

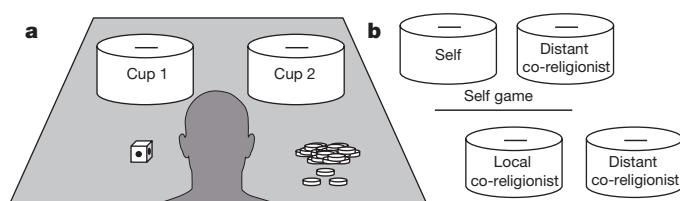


Figure 1 | The random allocation game. a, b, Generic game setup (a) and variants used in present work (b).

Table 2 | Log odds ratios for predicting allocations to distant co-religionists with 95% confidence intervals from our main binomial logistic regression models

Variable	Local co-religionist game	Local co-religionist game	Self game	Self game
Moralistic gods' punishment	1.15 (1.03, 1.27)*	—	1.11 (1.00, 1.23)†	—
Moralistic gods' knowledge	1.17 (1.00, 1.36)*	—	1.22 (1.05, 1.42)*	—
Punishment-knowledge combined	—	1.20 (1.04, 1.37)*	—	1.23 (1.07, 1.41)**
Material insecurity	1.02 (0.92, 1.12)	1.01 (0.92, 1.11)	0.98 (0.89, 1.08)	0.96 (0.88, 1.06)
Number of children	0.98 (0.96, 1.00)†	0.99 (0.97, 1.01)	0.98 (0.96, 1.00)*	0.98 (0.96, 1.00)*
<i>n</i>	503	519	504	520
Observations	15,090	15,570	15,120	15,600

All models in this table include field site and additional control variables as fixed effects. For all punishment-knowledge aggregate models, see Extended Data Table 1 (highlights from models 1 and 4 presented here) and Supplementary Table S9. See Supplementary Tables S5 and S6 for all other models (highlights from models 1 FE presented here). Odds ratios >1 indicate greater odds in allocating a coin to the distant co-religionist. ** $P \leq 0.01$, * $P \leq 0.05$, † $P \leq 0.10$.

s.d. = 3.31). One way to estimate the magnitude of these effects is to calculate the quotient of deviations from the ideal impartial allocation of 15. Compared to those who don't know, claiming the moralizing god punishes increases allocations towards distant co-religionists in the self game by a factor of 4.8 and in the local co-religionist game by a factor of 5.3. Extended Data Figs 4 and 5 detail the overall allocation distributions for both games.

We explored this relationship in more detail by regressing the number of coins allocated to the distant co-religionist cup on a host of variables for each game in a large set of binomial regressions (Extended Data Table 1 and Supplementary Information section S6). Table 2 shows a subset of the key predictors for the models with the largest set of control variables, including a number of economic and demographic variables such as education, material insecurity, number of children and field site fixed effects. Using sites as fixed effects allows us to remove the variation between our sites, so the results in Table 2 only capture the differences among individuals within sites. Based on previous work^{9,29}, we suspected that material insecurity and number of children would increase self and local favouritism, and therefore

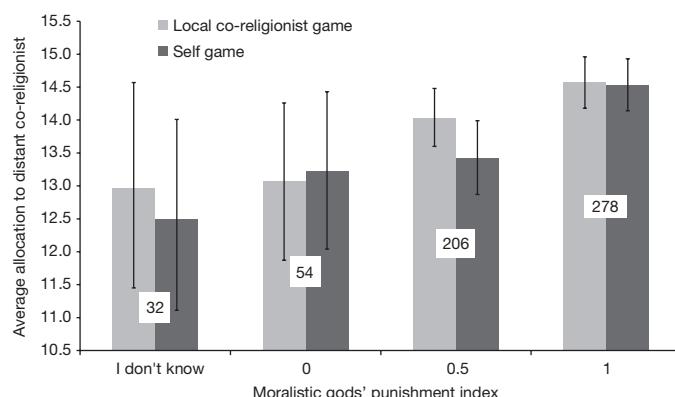


Figure 2 | Allocations to distant co-religionists increase as a function of moralistic gods' punishment. Punishment indices are mean values of a two-item scale (see Supplementary Information section S2.3.2). Error bars represent bootstrapped (1,000 replications) 95% confidence intervals of the mean. Histogram labels are sample sizes per category. Note that among the 32 individuals who responded "I don't know" to the questions pertaining to moralistic gods' punishment, 17 were Hadza and 15 were inland Tannese.

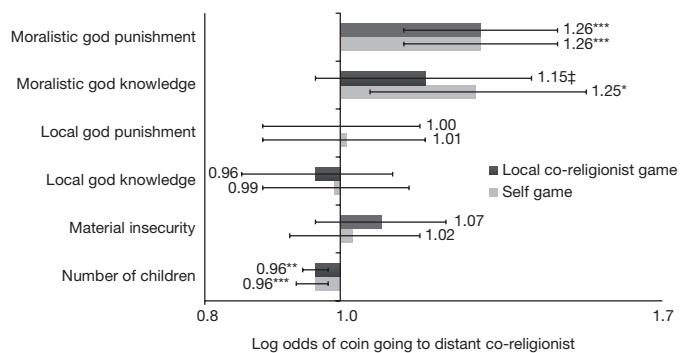


Figure 3 | Log odds ratios with 95% confidence interval plots of the influence of key variables on the odds that a coin goes into the cup for the distant co-religionist. Odds ratios >1 indicate an increase in odds and odds ratios <1 indicate a decrease in the odds of allocating a coin to the distant co-religionist (** $P \leq 0.001$, ** $P \leq 0.01$; * $P \leq 0.05$; † $P \leq 0.15$). The x axis is on a logarithmic scale. Both models include other controls ($n = 390$). Local co-religionist and self results include sites as fixed effects. Note that Indo-Fijians are not included in these models due to the lack of data for local gods. See Supplementary Tables S5 and S6 for full models (models 2FE are presented here).

we include both in our model (Supplementary Information section S2.3.1). To affirm the robustness of these analyses, we estimated many alternative models, formulated mixed models, and used both alternative standard error estimates and different approaches to modelling the error (Supplementary Information section S5.4). Across a wide range of specifications and models including a host of variables (for example, divine rewards, emotional closeness to distant co-religionists, among others), both moralistic gods' punishment and knowledge, as well as our aggregate punishment-knowledge variable, are reliably associated with less bias against distant co-religionists (Supplementary Tables S5–S9).

We checked whether the effects of moralistic gods' punishment and knowledge were indeed specific to powerful, moralizing gods. We added local gods' punishment and knowledge to the models presented in Table 2. Figure 3 shows the odds ratios and confidence intervals for these coefficients. Although neither the punishing powers nor knowledge of these local deities had any association with the allocations, the odds ratios for our key predictors pertaining to moralistic gods actually increased. These overall findings are correlational and should be interpreted with caution and in combination with other evidence, also considering that religious priming did not reveal consistent effects. However, these patterns reduce concerns that omitted third variables might account for the correlations we observe. A third variable, in addition to correlating with allocations, would have to correlate only with the punishing and knowing character of moralistic and knowledgeable gods, but not with those same attributes in local gods or with the tendency of either type of deity to reward people.

These results build on previous findings and have important implications for understanding the evolution of the wide-ranging cooperation found in large-scale societies. Moreover, when people are more inclined to behave impartially towards others, they are more likely to share beliefs and behaviours that foster the development of larger-scale cooperative institutions, trade, markets and alliances with strangers. This helps to partly explain two phenomena: the evolution of large and complex human societies and the religious features of societies with greater social complexity that are heavily populated by such gods^{6,7}. In addition to some forms of religious rituals and non-religious norms and institutions, such as courts, markets and police, the present results point to the role that commitment to knowledgeable, moralistic and punitive gods plays in solidifying the social bonds that create broader imagined communities^{11,12,31}.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 June 2015; accepted 7 January 2016.

Published online 10 February 2016.

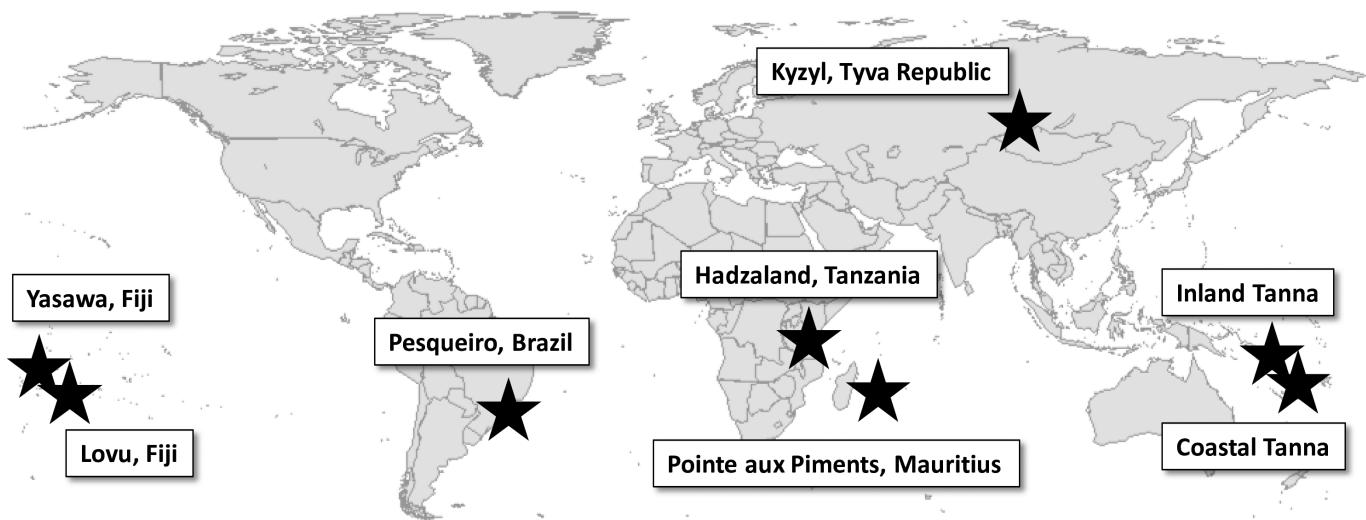
1. Richerson, P. J. & Boyd, R. Complex societies: the evolutionary origins of a crude superorganism. *Hum. Nat.* **10**, 253–289 (1999).
2. Turchin, P. in *Cultural Evolution: Society, Technology, Language, and Religion* (eds Richerson, P. J. & Christiansen, M. H.) 61–73 (MIT Press, 2013).
3. Fehr, E., Fischbacher, U. & Gächter, S. Strong reciprocity, human cooperation, and the enforcement of social norms. *Hum. Nat.* **13**, 1–25 (2002).
4. Turchin, P., Currie, T. E., Turner, E. A. L. & Gavrilets, S. War, space, and the evolution of Old World complex societies. *Proc. Natl. Acad. Sci. USA* **110**, 16384–16389 (2013).
5. Johnson, D. D. P. God's punishment and public goods. *Hum. Nat.* **16**, 410–446 (2005).
6. Norenzayan, A. *Big Gods: How Religion Transformed Cooperation and Conflict* (Princeton Univ. Press, 2013).
7. Norenzayan, A. et al. The cultural evolution of prosocial religions. *Behav. Brain Sci.* <http://dx.doi.org/10.1017/S0140525X14001356> (2015).
8. Schloss, J. P. & Murray, M. J. Evolutionary accounts of belief in supernatural punishment: a critical review. *Relig. Brain Behav.* **1**, 46–99 (2011).
9. McNamara, R. A., Norenzayan, A. & Henrich, J. Supernatural punishment, in-group biases, and material insecurity: experiments and ethnography from Yasawa, Fiji. *Relig. Brain Behav.* **6**, 34–55 (2016).
10. Rossano, M. J. Supernaturalizing social life. *Hum. Nat.* **18**, 272–294 (2007).
11. Sosis, R. & Bressler, E. R. Cooperation and commune longevity: a test of the costly signaling theory of religion. *Cross-Cultural Res.* **37**, 211–239 (2003).
12. Sosis, R. & Ruffle, B. J. Religious ritual and cooperation: testing for a relationship on Israeli religious and secular kibbutzim. *Curr. Anthropol.* **44**, 713–722 (2003).
13. Atran, S. & Henrich, J. The evolution of religion: how cognitive by-products, adaptive learning heuristics, ritual displays, and group competition generate deep commitments to prosocial religions. *Biol. Theory* **5**, 18–30 (2010).
14. Pew Research Centres *The Future of World Religions: Population Growth Projections 2010–2050*. http://www.pewforum.org/files/2015/03/PF_15.04.02_ProjectionsFullReport.pdf (2015).
15. Botero, C. A. et al. The ecology of religious beliefs. *Proc. Natl. Acad. Sci. USA* **111**, 16784–16789 (2014).
16. Watts, J. et al. Broad supernatural punishment but not moralizing high gods precede the evolution of political complexity in Austronesia. *Proc. R. Soc. Lond. B* **282**, 20142556 (2015).
17. Haidt, J. & Kesebir, S. in *Handbook of Social Psychology* 797–832 (Wiley, 2010).
18. Purzycki, B. G. The minds of gods: a comparative study of supernatural agency. *Cognition* **129**, 163–179 (2013).
19. Purzycki, B. G. *Tyan cher eazi* and the socioecological constraints of supernatural agents' minds. *Relig. Brain Behav.* **1**, 31–45 (2011).
20. Purzycki, B. G. et al. What does God know? Supernatural agents' access to socially strategic and non-strategic information. *Cogn. Sci.* **36**, 846–869 (2012).
21. Atkinson, Q. D. & Bourrat, P. Beliefs about God, the afterlife and morality support the role of supernatural policing in human cooperation. *Evol. Hum. Behav.* **32**, 41–49 (2011).
22. Shariff, A. F. & Hemtulla, M. Divergent effects of beliefs in heaven and hell on national crime rates. *PLoS ONE* **7**, e39048 (2012).
23. Bering, J. M., McLeod, K. & Shackelford, T. K. Reasoning about dead agents reveals possible adaptive trends. *Hum. Nat.* **16**, 360–381 (2005).
24. Piazza, J., Bering, J. M. & Ingram, G. 'Princess Alice is watching you': children's belief in an invisible person inhibits cheating. *J. Exp. Child Psychol.* **109**, 311–320 (2011).
25. Shariff, A. F., Willard, A. K., Andersen, T. & Norenzayan, A. Religious priming: a meta-analysis with a focus on prosociality. *Personal. Soc. Psychol. Rev.* **20**, 27–48 (2016).
26. Henrich, J., Heine, S. J. & Norenzayan, A. The weirdest people in the world? *Behav. Brain Sci.* **33**, 61–83 (2010).
27. Henrich, J. et al. Markets, religion, community size, and the evolution of fairness and punishment. *Science* **327**, 1480–1484 (2010).
28. Cohn, A., Fehr, E. & Maréchal, M. A. Business culture and dishonesty in the banking industry. *Nature* **516**, 86–89 (2014).
29. Hruschka, D. et al. Impartial institutions, pathogen stress and the expanding social network. *Hum. Nat.* **25**, 567–579 (2014).
30. Chuah, S.-H., Hoffmann, R., Ramasamy, B. & Tan, J. H. W. Religion, ethnicity and cooperation: an experimental study. *J. Econ. Psychol.* **45**, 33–43 (2014).
31. Xygalatas, D. et al. Extreme rituals promote prosociality. *Psychol. Sci.* **24**, 1602–1605 (2013).

Supplementary Information is available in the online version of the paper.

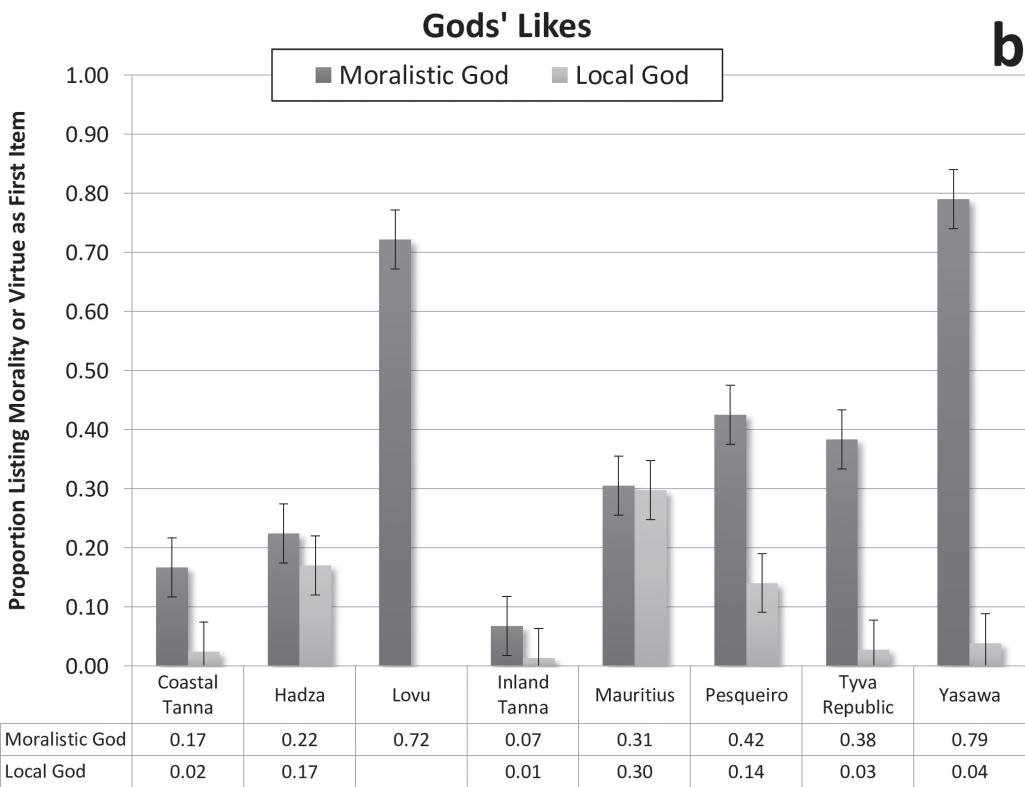
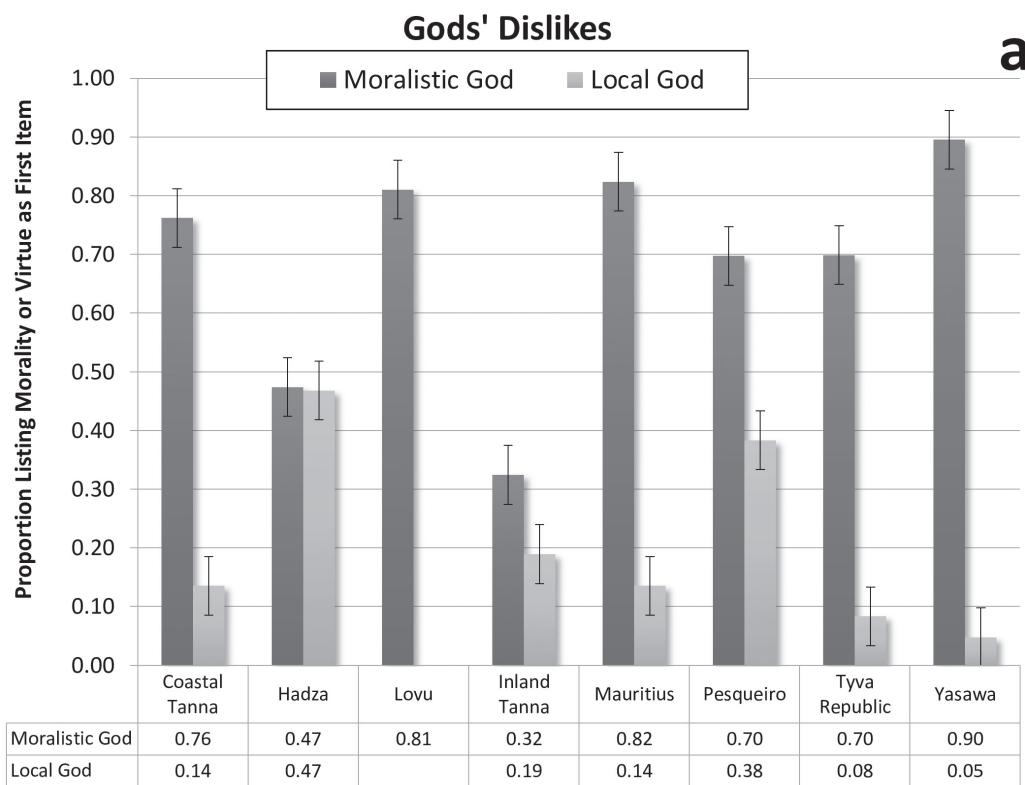
Acknowledgements We thank everyone who participated in the study and our local assistants without whom this project would not have been possible. We acknowledge funding from a research grant, "The Emergence of Prosocial Religions" from the John Templeton Foundation, and the Cultural Evolution of Religion Research Consortium (CERC), funded by a generous partnership grant (895–2011–1009) from the Social Sciences and Humanities Research Council of Canada. Q.D.A. is grateful for the support provided by a Rutherford Discovery Fellowship. E.C. thanks the British Academy for Fellowship support, J.H. thanks the Canadian Institute for Advanced Research for support, A.N. thanks the James McKeen Cattell Foundation for sabbatical support, and B.G.P. thanks L. Loveridge and J. McCutcheon. We thank A. Baimel, A. Barnett, J. Bulbulia, N. Chan, M. Collard, T. Lai, J. Lanman, B. Milner, M. Muthukrishna, C. Placek, E. Slingerland, R. Sosis, H. Whitehouse and C. Xu.

Author Contributions J.H., A.N. and B.G.P. conceived the study, prepared protocols, managed project communication and initiated manuscript preparation. C.A., Q.D.A., E.C., R.A.M., A.K.W., B.G.P. and D.X. collected data. B.G.P. conducted all analyses, made all graphs, Tables and illustrations. All authors participated in developing and refining protocols, experimental design and in manuscript preparation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.G.P. (bgpurzycki@alumni.ubc.ca) or J.H. (henrich@psych.ubc.ca).



Extended Data Figure 1 | Map of the eight field site locations. Map from R package 'maps' (2015). R version by Ray Brownrigg. R package version 3.0.0-2 (<http://CRAN.R-project.org/package=maps>).

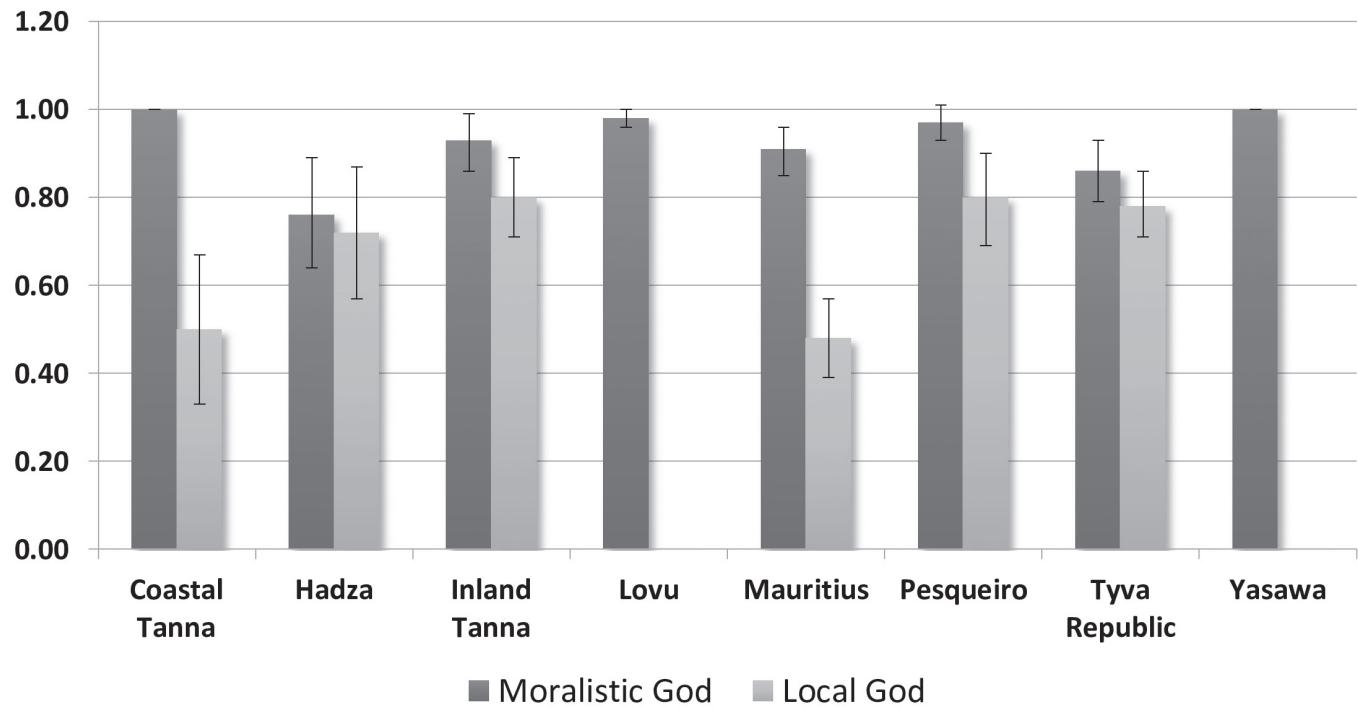


Extended Data Figure 2 | Proportion of sample listing moral and virtue items for moralistic and local gods' dislikes and likes by site. a, b, We asked participants to freely list up to five things that moralistic and local gods dislike and like. These free-list items were subsequently coded by two independent coders using 12 categories (see Supplementary Information

section S4.1.1 for the methods). Items listed first are the most salient items in participants' models of gods' concerns. Error bars have a total breadth of 10%. Note that Indo-Fijians (Lovu) did not answer questions about local gods.

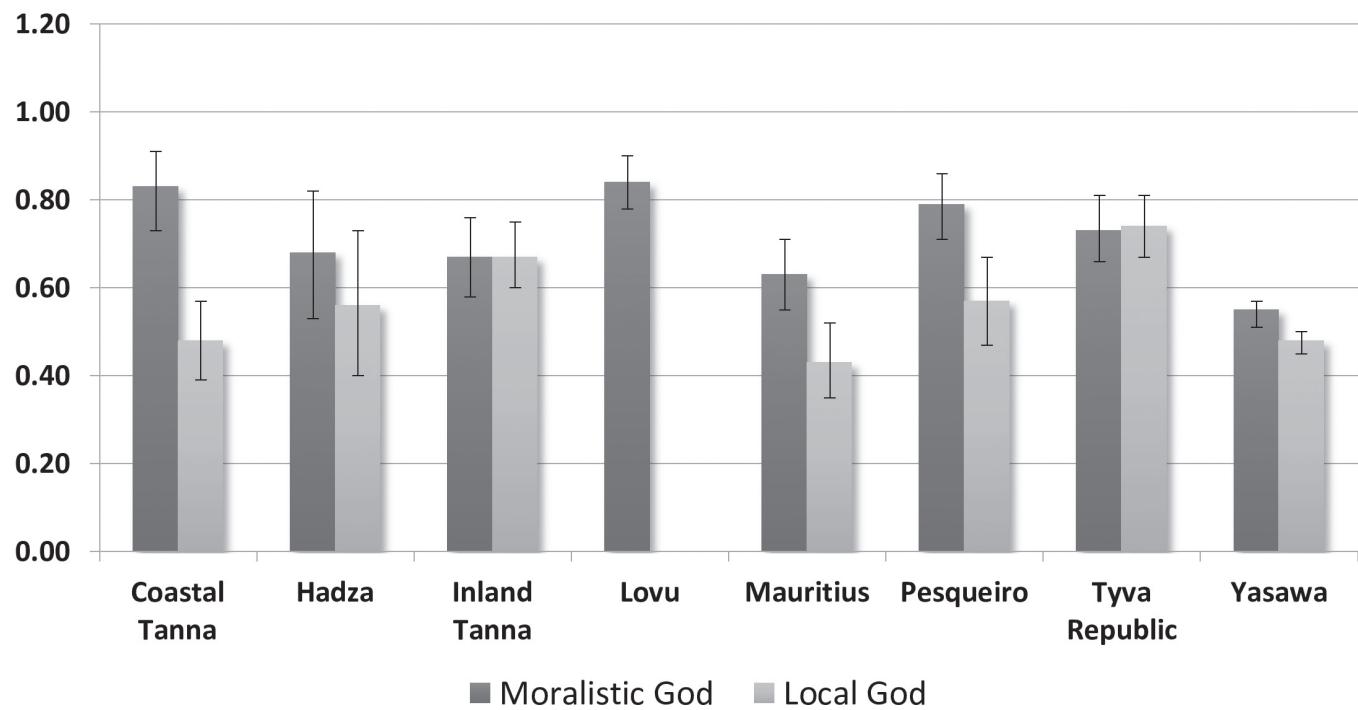
Knowledge Scale

a



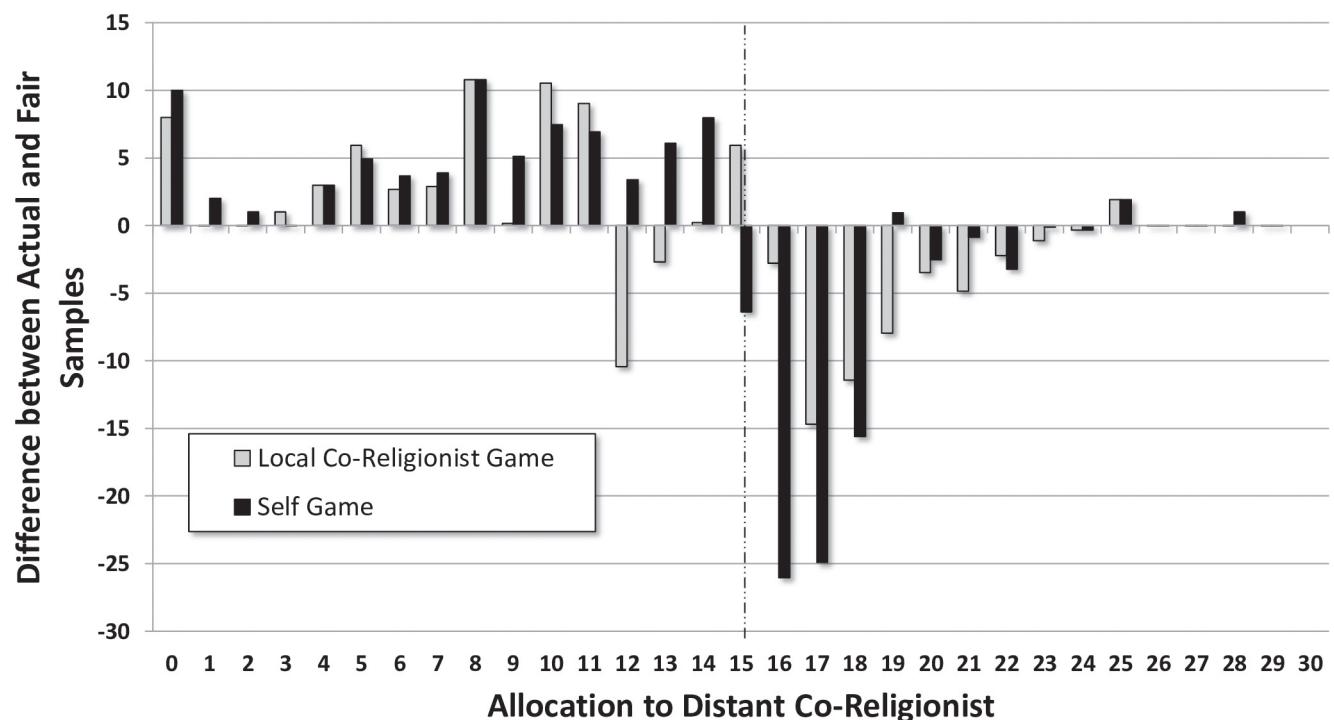
Punishment Scale

b



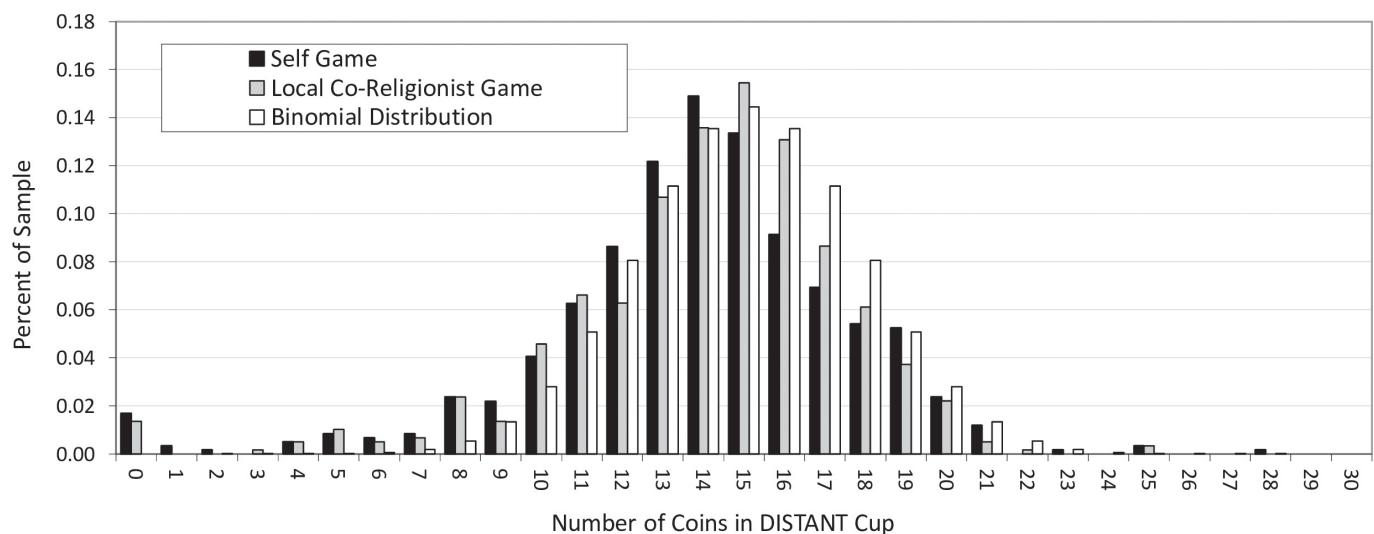
Extended Data Figure 3 | Mean moralistic and local gods' knowledge and punishment scales by site. a, b, Error bars represent 95% confidence intervals of the mean. Lovu (Indo-Fijians) did not answer questions about local gods, and Yasawans' (native Fijians) attributions of ancestor spirits'

knowledge had a mean and standard deviation of zero. Note that local gods often punish for non-moralistic reasons. See Supplementary Information sections S4.1 and S4.2 for methods and analyses.



Extended Data Figure 4 | Plot of differences between size of actual allocations and allocations from binomially distributed sample of the same size. The halfway mark of 15 indicates the predicted mean of all cups. Bars above zero on the y axis indicate higher frequencies of

allocations than predicted, and those lower indicate fewer individuals than predicted. Note the cluster of extremely lower-than-predicted values immediately after the cut-off point of 15.



Extended Data Figure 5 | Per cent of sample by allocation amount to distant cup in local co-religionist (grey) and self games (black) as compared to binomial distribution (white). For both games, allocations lean towards the left of a theoretically ideal binomial distribution

suggesting that overall, participants biased allocations towards themselves ($n = 591$) and local co-religionists ($n = 589$) at the expense of geographically distant co-religionists.

Extended Data Table 1 | Models accounting for allocations to anonymous distant co-religionists with punishment–knowledge aggregate scales

	1 Local Co-Religionist Game	2 Local Co-Religionist Game	3 Local Co-Religionist Game	4 Self Game	5 Self Game	6 Self Game
Punishment-Knowledge (MG)	1.20 [1.04, 1.37]*	1.27 [1.08, 1.50]**	1.47 [1.18, 1.83]***	1.23 [1.07, 1.41]**	1.40 [1.18, 1.65]***	1.51 [1.21, 1.88]***
Punishment-Knowledge (LG)	--	0.95 [0.83, 1.09]	0.90 [0.76, 1.07]	--	1.00 [0.88, 1.14]	1.04 [0.88, 1.23]
Reward (MG)	--	--	1.03 [0.86, 1.24]	--	--	1.10 [0.92, 1.33]
Reward (LG)	--	--	0.92 [0.79, 1.08]	--	--	0.93 [0.79, 1.09]
Moral Index (MG)	--	--	0.98 [0.93, 1.03]	--	--	1.00 [0.95, 1.05]
Moral Index (LG)	--	--	1.03 [0.98, 1.08]	--	--	0.98 [0.93, 1.03]
Age (Centered)	1.00 [1.00, 1.00]	1.00 [1.00, 1.01]*	1.01 [1.00, 1.01]**	1.00 [1.00, 1.01]*	1.01 [1.00, 1.01]***	1.01 [1.00, 1.01]***
Sex (0 = Female)	1.06 [0.99, 1.14]†	1.05 [0.97, 1.13]	1.09 [1.00, 1.19]†	1.03 [0.96, 1.10]	1.07 [0.99, 1.15]†	1.06 [0.97, 1.15]
Years of Formal Education	0.99 [0.98, 1.00]	0.99 [0.98, 1.01]	1.00 [0.98, 1.01]	1.00 [0.99, 1.01]	1.00 [0.99, 1.02]	1.01 [0.99, 1.02]
Material Insecurity	1.01 [0.92, 1.11]	1.08 [0.97, 1.20]	1.09 [0.96, 1.23]‡	0.96 [0.88, 1.06]	1.03 [0.92, 1.15]	1.01 [0.89, 1.14]
# of Children	0.99 [0.97, 1.01]	0.97 [0.95, 0.99]**	0.97 [0.94, 0.99]*	0.98 [0.96, 1.00]*	0.96 [0.94, 0.98]***	0.96 [0.93, 0.99]**
DISTANT Religious Similarity	--	--	1.01 [0.98, 1.05]	--	--	1.03 [0.98, 1.08]
LOCAL Emotional Closeness	--	--	0.98 [0.94, 1.03]	--	--	1.00 [0.95, 1.04]
DISTANT Emotional Closeness	--	--	0.97 [0.94, 1.01]	--	--	1.00 [0.97, 1.04]
Police Evaluation	--	--	1.03 [0.98, 1.09]	--	--	1.03 [0.98, 1.09]
Honesty Mentioned (0 = No)	--	--	0.92 [0.76, 1.11]	--	--	1.19 [0.99, 1.44]†
Treatment (0 = Control)	1.00 [0.93, 1.08]	1.04 [0.96, 1.14]	1.05 [0.95, 1.15]	0.97 [0.90, 1.05]	0.96 [0.89, 1.05]	0.95 [0.87, 1.04]
Game Order (0 = Self First)	1.05 [0.99, 1.12]‡	1.08 [1.01, 1.16]*	1.12 [1.03, 1.22]**	1.03 [0.96, 1.09]	1.06 [0.98, 1.14]‡	1.11 [1.02, 1.20]*
Coastal Tanna	0.98 [0.83, 1.16]	1.06 [0.86, 1.30]	1.08 [0.86, 1.34]	0.87 [0.73, 1.02]†	0.89 [0.73, 1.10]	0.89 [0.71, 1.11]
Hadza	0.59 [0.48, 0.73]***	0.70 [0.55, 0.89]**	--	0.63 [0.51, 0.78]***	0.76 [0.59, 0.97]*	--
Inland Tanna	0.92 [0.75, 1.14]	0.98 [0.77, 1.24]	1.02 [0.77, 1.35]	0.80 [0.65, 0.99]*	0.89 [0.70, 1.12]	0.89 [0.67, 1.18]
Lovu	0.96 [0.83, 1.11]	--	--	0.91 [0.79, 1.06]	--	--
Mauritius	0.90 [0.78, 1.04]‡	0.88 [0.76, 1.03]‡	0.91 [0.75, 1.10]	0.77 [0.66, 0.88]***	0.78 [0.67, 0.91]**	0.78 [0.65, 0.94]*
Pesqueiro	0.91 [0.78, 1.06]	0.91 [0.77, 1.08]	0.89 [0.72, 1.10]	0.90 [0.77, 1.04]	0.84 [0.71, 0.99]*	0.94 [0.77, 1.16]
Yasawa	0.79 [0.69, 0.91]**	0.77 [0.66, 0.90]**	0.80 [0.60, 1.06]‡	0.62 [0.54, 0.72]***	0.63 [0.54, 0.74]***	0.64 [0.48, 0.86]**
Constant	0.88 [0.70, 1.12]	0.86 [0.66, 1.13]	0.85 [0.60, 1.22]	0.98 [0.78, 1.24]	0.81 [0.61, 1.06]‡	0.68 [0.47, 0.97]*
AIC	2846.5	2224.2	1770.3	2990.4	2351.1	1920.7
Pseudo- R^2	0.19	0.18	0.19	0.23	0.26	0.31
N	519	408 ^a	323 ^b	520	408 ^a	323 ^b

Log odds ratios (and 95% CIs) for variables predicting allocations to the distant co-religionist.

^aLovu not included.

^bHadza and Lovu not included.

All models are binomial logistic regressions, backward selected for site inclusion. Models include field sites as fixed effects. Moralistic god variables are denoted as 'MG' and local god variables are denoted as 'LG'. Pseudo R^2 s are Nagelkerke's R^2 . *** $P \leq 0.001$, ** $P \leq 0.01$, * $P \leq 0.05$, † $P \leq 0.10$, ‡ $P \leq 0.15$. See Supplementary Information section S2.3 for variable definitions, Supplementary Information section S5 for discussion of analyses and Supplementary Table S9 for punishment–knowledge aggregate models with extreme values removed.

The genome of the seagrass *Zostera marina* reveals angiosperm adaptation to the sea

Jeanine L. Olsen^{1*}, Pierre Rouzé², Bram Verhelst², Yao-Cheng Lin², Till Bayer³, Jonas Collen⁴, Emanuela Dattolo⁵, Emanuele De Paoli⁶, Simon Dittami⁴, Florian Maumus⁷, Gurvan Michel⁴, Anna Kersting^{8,9}, Chiara Lauritano⁵, Rolf Lohaus², Mats Töpel¹⁰, Thierry Tonon⁴, Kevin Vanneste², Mojgan Amirebrahimi¹¹, Janina Brakel³, Christoffer Boström¹², Mansi Chovatia¹¹, Jane Grimwood^{11,13}, Jerry W. Jenkins^{11,13}, Alexander Jueterbock¹⁴, Amy Mraz¹⁵, Wytze T. Stam¹, Hope Tice¹¹, Erich Bornberg-Bauer⁸, Pamela J. Green¹⁶, Gareth A. Pearson¹⁷, Gabriele Procaccini^{5*}, Carlos M. Duarte¹⁸, Jeremy Schmutz^{11,13}, Thorsten B. H. Reusch^{3,19*} & Yves Van de Peer^{2,20,21*}

Seagrasses colonized the sea¹ on at least three independent occasions to form the basis of one of the most productive and widespread coastal ecosystems on the planet². Here we report the genome of *Zostera marina* (L.), the first, to our knowledge, marine angiosperm to be fully sequenced. This reveals unique insights into the genomic losses and gains involved in achieving the structural and physiological adaptations required for its marine lifestyle, arguably the most severe habitat shift ever accomplished by flowering plants. Key angiosperm innovations that were lost include the entire repertoire of stomatal genes³, genes involved in the synthesis of terpenoids and ethylene signalling, and genes for ultraviolet protection and phytochromes for far-red sensing. Seagrasses have also regained functions enabling them to adjust to full salinity. Their cell walls contain all of the polysaccharides typical of land plants, but also contain polyanionic, low-methylated pectins and sulfated galactans, a feature shared with the cell walls of all macroalgae⁴ and that is important for ion homoeostasis, nutrient uptake and O₂/CO₂ exchange through leaf epidermal cells. The *Z. marina* genome resource will markedly advance a wide range of functional ecological studies from adaptation of marine ecosystems under climate warming^{5,6}, to unravelling the mechanisms of osmoregulation under high salinities that may further inform our understanding of the evolution of salt tolerance in crop plants⁷.

Seagrasses are a polyphyletic assemblage of basal monocots belonging to four families in the Alismatales^{1,2} (Supplementary Note 1.1 and Supplementary Fig. 1.1). As a functional group, they provide the foundation of highly productive ecosystems present along the coasts of all continents except Antarctica, where they rival tropical rain forests and coral reefs in ecosystem services^{8,9}. In colonizing sedimentary shorelines of the world's ocean, seagrasses found a vast new habitat free of terrestrial competitors and insect pests but had to adapt to cope with new structural and physiological challenges related to full marine conditions.

Zostera marina (Zosteraceae), or eelgrass (Fig. 1), is the most widespread species throughout the temperate northern hemisphere of

the Pacific and Atlantic¹⁰. A clone of *Z. marina* was sequenced from the Archipelago Sea, southwest Finland, using a combination of fosmid-ends and whole-genome shotgun (WGS) approaches (Methods, Supplementary Note 2). The 202.3 Mb *Z. marina* genome encodes 20,450 protein-coding genes, 86.6% of which (17,511 genes, Supplementary Note 3.1) are supported by transcriptome data from leaves, roots and flowers (Extended Data Fig. 1, Supplementary Notes 3.2–3.3 and Supplementary Data 1–3). Genes are located in numerous gene-dense islands separated by stretches of repeat elements accounting for 63% of the non-gapped assembly (Extended Data Fig. 2, Supplementary Note 3.1) as compared to only 13% in the only other sequenced alismatid, the freshwater duckweed, *Spirodela polyrhiza* (Alismatales, Araceae)¹¹. Gypsy-type (32%) and Copia-type (20%) transposable elements contribute to most of the repetitive DNA. Sequence divergence analysis suggests that the genome retains copies from two distinct periods of invasion by Copia elements, but only one period for Gypsy elements (Extended Data Fig. 3a–c). Genes gained by *Z. marina* ('accessory') are located closer to transposable elements than to conserved ('single copy') genes (Fisher's exact test, $P < 0.0001$) indicating that transposable elements may have played a role in genic adaptation.

We identified 36 conserved microRNAs with high confidence and their predicted targets (Supplementary Note 3.4, Supplementary Data 4 and 5). A novel variant of miR528 (not present in *Spirodela*) was found to be the only member of this miRNA family, and demonstrates that this conserved miRNA is the only one ancestral to the entire monocot lineage. Most likely, *Z. marina* did not take part in the subsequent birth of miRNAs that are common to several other monocots¹²; nor did it experience or retain traces of prominent miRNA duplications.

Analysis of synonymous substitutions per synonymous site (K_s) age distributions indicates that *Z. marina* carries the remnants of an independent, ancient whole-genome duplication (WGD) event (Fig. 2a, Supplementary Note 4.1)¹³. Duplicated segments account for ~9% of the *Z. marina* genome, probably an underestimate due to the fragmented nature of the assembly. *Zostera* and *Spirodela* diverged somewhere between 135 and 107 million years ago (Mya)¹⁴ and

¹Groningen Institute of Evolutionary Life Sciences (GELIFES), University of Groningen, PO Box 11103, 9700 CC Groningen, The Netherlands. ²Department of Plant Systems Biology, VIB and Department of Plant Biotechnology and Bioinformatics, Ghent University, Technologiepark 927, B-9052 Ghent, Belgium. ³GEOMAR Helmholtz Centre for Ocean Research-Kiel, Evolutionary Ecology, Düsternbrooker Weg 20, D-24105 Kiel, Germany. ⁴Sorbonne Université, UPMC Univ Paris 06, CNRS, UMR 8227, Integrative Biology of Marine Models, Station Biologique de Roscoff, CS 90074, F-29688, Roscoff cedex, France. ⁵Stazione Zoologica Anton Dohrn, Villa Comunale, 80121 Naples, Italy. ⁶Dipartimento di Scienze Agrarie e Ambientali, University of Udine, Via delle Scienze 206, 33100 Udine, Italy. ⁷INRA, UR1164 URGI—Research Unit in Genomics-Info, INRA de Versailles-Grignon, Route de Saint-Cyr, Versailles 78026, France. ⁸Institute for Evolution and Biodiversity, Westfälische Wilhelms-University of Münster, Hüfferstrasse 1, D-48149 Münster, Germany. ⁹Institute for Computer Science, Heinrich Heine University, D-40255 Duesseldorf, Germany. ¹⁰Department of Biological and Environmental Sciences, Bioinformatics Infrastructure for Life Sciences (BILS), University of Gothenburg, Medicinaregatan 18A, 40530 Gothenburg, Sweden.

¹¹Department of Energy Joint Genome Institute, 2800 Mitchell Dr., #100, Walnut Creek, California 94598, USA. ¹²Environmental and Marine Biology, Faculty of Science and Engineering, Åbo Akademi University, Artillerigatan 6, FI-20520 Turku/Åbo, Finland. ¹³HudsonAlpha Institute for Biotechnology, 601 Genome Way NW, Huntsville, Alabama 35806, USA. ¹⁴Marine Ecology Group, Nord University, Postbox 1490, 8049 Bodø, Norway. ¹⁵Amplicon Express, 2345 NE Hopkins Ct., Pullman, Washington 99163, USA. ¹⁶School of Marine Science and Policy, Department of Plant and Soil Sciences, Delaware Biotechnology Institute, University of Delaware, 15-Innovation Way, Newark, Delaware 19711, USA. ¹⁷Marine Ecology and Evolution, Centre for Marine Sciences (CCMAR), University of Algarve, 8005-139 Faro, Portugal. ¹⁸King Abdullah University of Science and Technology (KAUST), Red Sea Research Center (RSRC), Thuwal 23955-6900, Saudi Arabia.

¹⁹University of Kiel, Faculty of Mathematics and Natural Sciences, Christian-Albrechts-Platz 4, 24118 Kiel, Germany. ²⁰Genomics Research Institute, University of Pretoria, Hatfield Campus, Pretoria 0028, South Africa. ²¹Bioinformatics Institute Ghent, Ghent University, Ghent B-9000, Belgium.

*These authors contributed equally to this work.

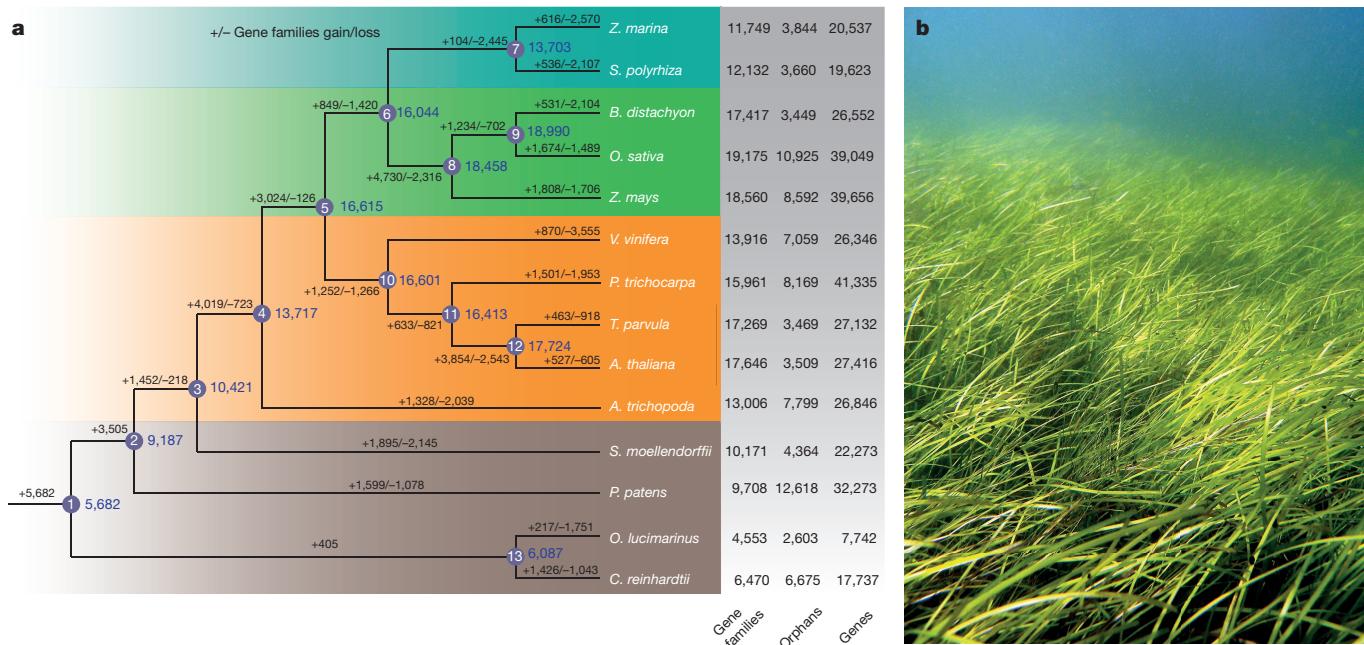


Figure 1 | Zostera marina and phylogenetic tree showing gene family expansion/contraction analysis compared with 13 representatives of the Viridiplanteae. a, Gains and losses are indicated along branches and nodes. The number of gene families, orphans (single-copy gene families) and

phylogenomic dating¹³ of the *Z. marina* WGD suggests that it occurred 72–64 Mya (Fig. 2b), thus independently from the two WGDs reported for *S. polystachya*¹¹. This timeframe coincides with the initial diversification of a freshwater clade that includes three of the four families of seagrasses (Supplementary Table 1.1) and with the Cretaceous–Palaeogene (K–Pg) extinction event (Fig. 2c), which provided new ecological opportunities and may have triggered seagrass adaptive radiations.

We mapped signatures of loss and gain of gene families (Supplementary Note 4.2) onto a phylogenetic tree (Fig. 1a). We also mapped losses and gains of Pfam domains (Supplementary Fig. 4.4,

number of predicted genes is indicated next to each species. Background colours (top to bottom) are Alismatales, other monocots, dicots, mosses/algae **b**, Typical *Zostera marina* meadow, Archipelago Sea, southwest Finland (photo by C.B.).

Supplementary Data 6). While many genes are shared between *Zostera* and *Spirodela*, clearly some losses and gains are unique to *Zostera* in relation to its marine environment, the alismatid lineage having set the stage for the subsequent freshwater–marine transition. Those unique to *Z. marina* include the absence of all the genes involved in stomatal differentiation (Fig. 3a, Extended Data Table 1 and Supplementary Note 5.1) and the disappearance of genes comprising entire pathways encoding volatiles synthesis and sensing (Supplementary Note 6.1), such as those for ethylene¹⁵ (Fig. 3b, Extended Data Table 2). Terpenoid genes are also drastically reduced to two (Fig. 3c), as compared with four in *Spirodela*, 50 in *Oryza* and >100 in *Eucalyptus*, thus

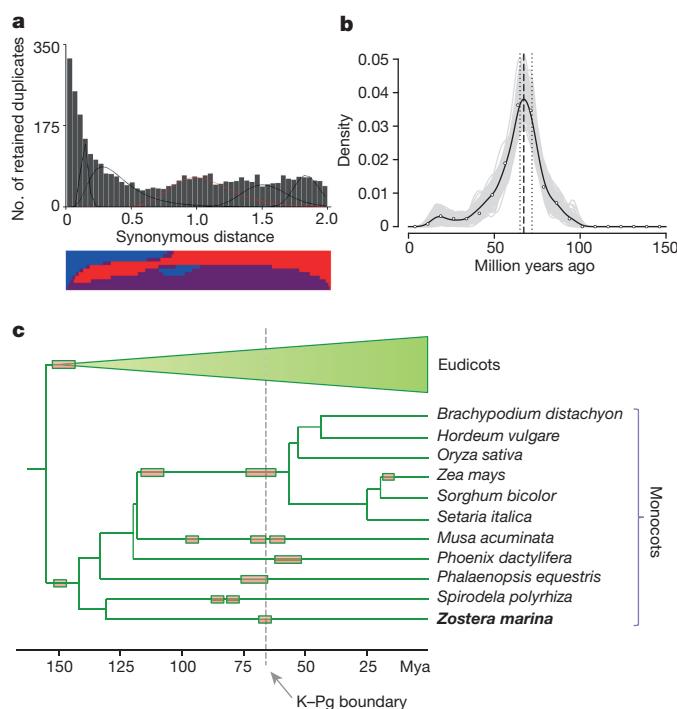


Figure 2 | Ancient whole-genome duplication (WGD). a, K_S -based age distribution of the whole *Z. marina* paralome. The x axis shows the synonymous distance until a K_S cut-off of 2, in bins of 0.04, containing the K_S values that were used for mixture modelling (excluding those with a $K_S \leq 0.1$). The component of the Gaussian mixture model plotted in red (as identified by EMMIX) corresponds to a WGD feature based on the SiZer analysis (other components are shown in black). The transition from the blue to the red at a K_S of ~0.8 in the SiZer panel (below) indicates a change in the distribution and therefore provides evidence for an ancient WGD (Supplementary Table 4.1, Supplementary Fig. 4.1). **b**, Absolute age distribution obtained by phylogenomic dating of *Z. marina* paralogues. The solid black line represents the kernel density estimate (KDE) of the dated paralogues and the vertical dashed black line represents its peak, used as the consensus WGD age estimate, at 67 Mya. Grey lines represent the density estimates from 2,500 bootstrap replicates and the vertical black dotted lines represent the corresponding 90% confidence interval for the WGD age estimate, 64–72 Mya. The original raw distribution of dated paralogues is indicated by the circles. The y axis represents the percentage of gene pairs. **c**, Pruned phylogenetic tree with indication of WGD events (boxes)²⁹. The Cretaceous–Palaeogene (K–Pg) boundary is indicated by an arrow.

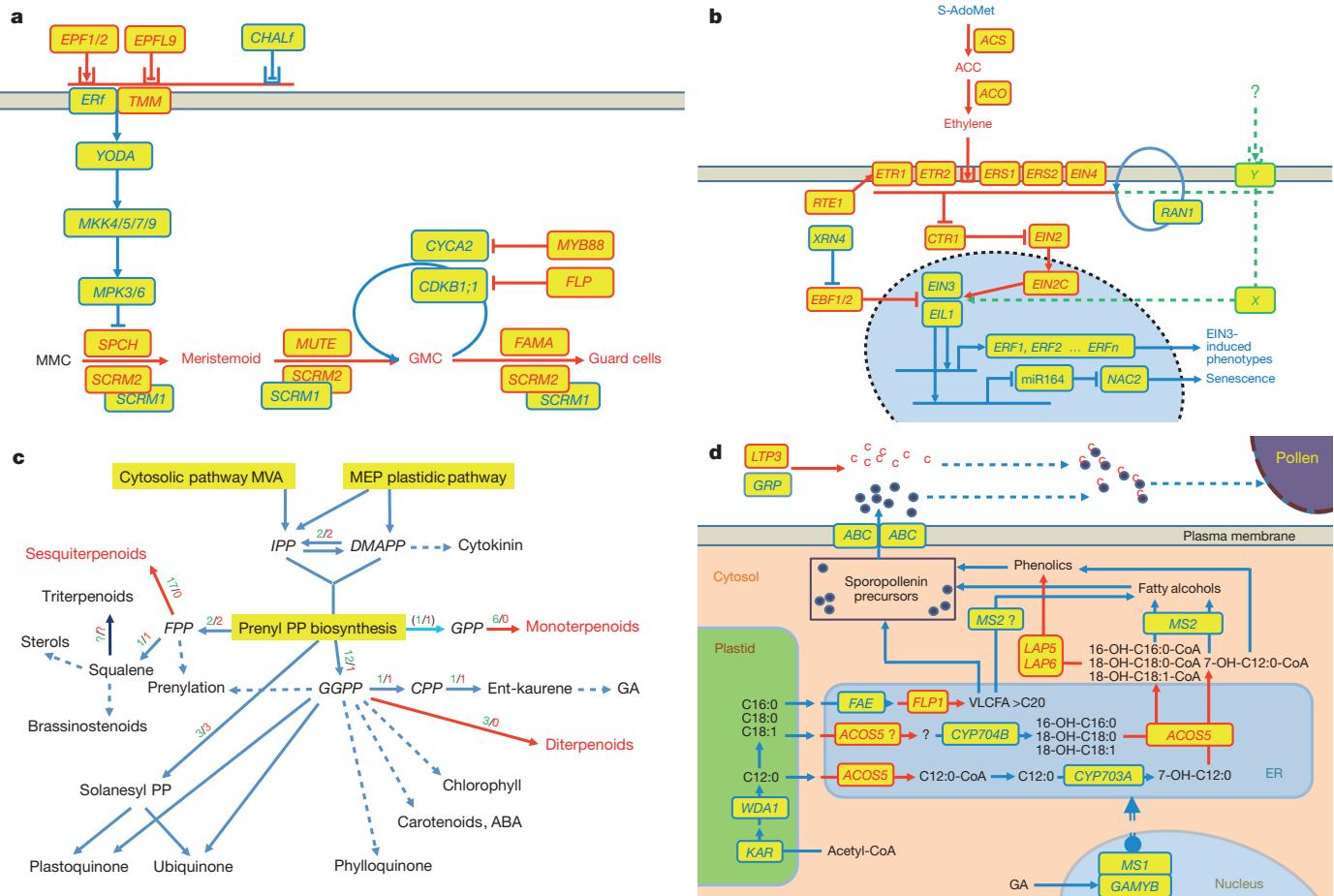


Figure 3 | Reconstruction of metabolic (or gene) pathways involved in the production of stomata, ethylene, terpene and pollen in *Z. marina*. **a**, Stomata differentiation from meristemoid mother cells (MMC) to guard mother cell (GMC) to guard cells. **b**, Ethylene synthesis and signalling up to *EIN2* have disappeared; *EIN3* and its downstream targets remain. **c**, Terpenoid biosynthesis in which the pathways producing volatiles are absent but those essential for primary metabolism remain. MVA, mevalonate; MEP, plastidic methylerythritol phosphate; IPP, isopentenyl pyrophosphate; DMAPP, dimethylallyl pyrophosphate; FPP, farnesyl pyrophosphate; GPP, geranyl diphosphate; GGPP, geranylgeranyl pyrophosphate; CPP, copalyd pyrophosphate; GA, gibberellic acid; PP, diphosphate; ABA, abscisic acid. **d**, Sporopollenin biosynthesis genes; regulatory genes in the nucleus control downstream processes (arrows) in response to signalling coming from external stimuli through receptors on the plasma membrane. All panels: genes in red are absent; blue are present; the grey line represents the plasma membrane. See Extended Data Tables 1–3.

isopentenyl pyrophosphate; DMAPP, dimethylallyl pyrophosphate; FPP, farnesyl pyrophosphate; GPP, geranyl diphosphate; GGPP, geranylgeranyl pyrophosphate; CPP, copalyd pyrophosphate; GA, gibberellic acid; PP, diphosphate; ABA, abscisic acid. **d**, Sporopollenin biosynthesis genes; regulatory genes in the nucleus control downstream processes (arrows) in response to signalling coming from external stimuli through receptors on the plasma membrane. All panels: genes in red are absent; blue are present; the grey line represents the plasma membrane. See Extended Data Tables 1–3.

precluding synthesis of secondary volatile terpenes (Supplementary Fig. 6.2). Only aromatic acid decarboxylases (AAAD) genes were expanded (Supplementary Fig. 6.3) and these form a clade distinct from *Spirodela*. The loss of volatiles is also consistent with the loss of stomata, through which they are emitted for airborne communication and plant defence. The repertoire of defence-related genes such as the six groups of NBS_LRR resistance genes (Supplementary Note 6.2) is also reduced to 44 (89 in *Spirodela* and 100–300 in other plants), which may be linked to a lower probability of infection of *Z. marina* due to the absence of stomata, which are a main entry point for pests and pathogens in terrestrial plants.

Land and aquatic floating plants (Embryophyta) are often exposed to intense ultraviolet (UV) radiation and have developed light sensing protein receptors with protective and signalling functions. In contrast, *Z. marina* inhabits a light-attenuated, submarine environment where it must cope with shifted spectral composition, characterized by low penetration of UV-B, red and far-red wavelengths¹⁶. Accordingly, *Z. marina* has lost ultraviolet-resistance (UVR8) genes associated with sensing and responding to UV damage (*Spirodela* has not), as well as phytochromes associated with red/far-red receptors (Supplementary Note 7). Whereas photosystems (PSI and PSII) are similar to those of other plants including *Spirodela*, members of the light-harvesting complex B (LHCB) family are expanded in number, possibly in combination

with non-photochemical quenching (NPQ), thereby enhancing performance at low light (Extended Data Fig. 4).

Seagrasses typically experience full marine seawater (35 g kg^{-1})¹⁷, whereas land plants obtain water with low osmolality ($0\text{--}2 \text{ g kg}^{-1}$) via the rhizosphere and aquatic plants experience fresh ($0\text{--}5 \text{ g kg}^{-1}$) to brackish ($0.5\text{--}20 \text{ g kg}^{-1}$) conditions. Although *Z. marina* displays a typical repertoire of Na^+ and K^+ antiporters (Supplementary Note 8, Supplementary Table 8.1), one of six H^+ -ATPase (AHA) genes (Supplementary Table 8.2, Supplementary Data 7) is strongly expressed in vegetative tissue and encodes a salt-tolerant H^+ -ATPase. Furthermore, *Z. marina* possesses three AHA genes (along with *Spirodela*) in a cluster unique to alismatids (Supplementary Fig. 8.1).

Uniquely, *Z. marina* has re-evolved new combinations of structural traits related to the cell wall. Synthesis of cutin-cuticular waxes to the outside of the leaf epidermis and suberin-lignin near the plasma membrane (Supplementary Note 9, Supplementary Table 9.1) surround a cell wall matrix of (hemi)celluloses, low-methylated pectin (zosterin) and macroalgal-like sulfated polysaccharides¹⁸ (Supplementary Note 10). The reduction in carbohydrate-related genes that modify the fine structure of cell wall hemicelluloses and pectins in *Z. marina* is not due to loss of pathways, but rather to the large variation within these CAZyme gene families in plants. Available genomes (including *Spirodela*) lack carbohydrate sulfotransferases and sulfatases, suggesting that land

plants have lost these genes as a key adaptation to terrestrial as well as freshwater conditions^{19,20}. In contrast, *Z. marina* has regained the ability to produce sulfated polysaccharides with an expansion of aryl sulfotransferases (12 genes) homologous to aryl sulfotransferases from land plants (Supplementary Note 10). Sulfation facilitates water and ion retention in the cell wall to cope with desiccation and osmotic stress at low tide and, likewise, low methylation of zosterin correlates with the expanded pectin carbohydrate esterase 8 (CE8) family, increasing the polyanionic character of the cell wall matrix. We speculate that several aryl sulfotransferases have evolved because carbohydrate sulfatases have been shown to be active on artificial aryl compounds such as methylumbelliferyl-sulfate²¹. Osmotic equilibrium is further achieved in *Z. marina* by organic osmolytes (mainly sucrose, trehalose and proline) in combination with a small cytoplasm:vacuole volume ratio (10%)²². Given that up to 90% of fixed carbon is stored as sucrose in the rhizomes, sucrose synthase (SuSy) and transport (SUT) genes are expanded while those for starch metabolism are greatly reduced, as expected in 'marine sugarcane' (Supplementary Note 7.2, Supplementary Data 8).

The repertoire of redox and other stress-resistance genes (Supplementary Note 8) is typical for angiosperms with the exception of catalase (CAT), which is reduced to a single copy in *Z. marina* (two in *Spirodela*). Late embryogenesis abundant (LEA) and dehydrins are clearly under-represented in both *Zostera* and *Spirodela* relative to other genomes. In contrast, *Zostera* possesses an unusual complement of metallothioneins. Aside from their role as chelators, metallothioneins may be involved in stress resistance; one of these, MT2L, is among the most highly constitutively expressed genes in *Z. marina* (Extended Data Fig. 5, Supplementary Note 8.2).

Sexual reproduction of *Z. marina* takes place underwater, involving completely submerged male and female flowers, and a unique exine-less, filiform pollen that winds around the bifurcate stigmas in a purely abiotic pollination process²³. Note that freshwater alismatids (and also *Spirodela*)²⁴ possess pollen with an exine layer. Exine-less pollen²⁵ is characteristic of all seagrasses except *Enhalus acoroides* (which is surface pollinated). Ten genes specifically involved in biosynthesis and modification of the pollen exine coat are missing; all other genes involved in the development of viable pollen remain intact (Fig. 3d, Extended Data Table 3, Supplementary Note 11.1). Finally, MADS-box gene transcription factors are also highly reduced to 50 in *Z. marina*, which is most likely related to its highly reduced flowers (also a feature of *Spirodela*) that lack the first two whorls of specialized floral leaves, calyx and corolla (Supplementary Note 11.2, Supplementary Table 11.2).

An increasing proportion of the world population inhabits the coastal zone. This impinges multiple pressures on ecosystems including seagrass beds^{26,27}, which in turn compromises the ecosystem services they may provide, including provisioning of harvestable fish and invertebrates, nutrient retention, carbon sequestration and erosion control. In the context of seagrass conservation, elucidating the genomic basis of *Z. marina*'s complex adaptations to ocean waters (Extended Data Fig. 6) will also inform the development of molecular indicators of their physiological status²⁸, as these unique ecosystems rank, unfortunately, among the most threatened on Earth^{26,27}.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 May; accepted 18 December 2015.

Published online 27 January 2016.

1. Les, D. H., Cleland, M. A. & Waycott, M. Phylogenetic studies in Alismatidae, II: evolution of marine angiosperms (seagrasses) and hydrophily. *Syst. Bot.* **22**, 443–463 (1997).
2. Larkum, W. D., Orth, R. J. & Duarte, C. M. *Seagrasses: Biology, Ecology and Conservation* (Springer, Dordrecht, Netherlands, 2006).
3. Berry, J. A., Beerling, D. J. & Franks, P. J. Stomata: key players in the earth system, past and present. *Curr. Opin. Plant Biol.* **13**, 232–239 (2010).

4. Aquino, R. S., Landeira-Fernandez, A. M., Valente, A. P., Andrade, L. R. & Mourao, P. A. S. Occurrence of sulfated galactans in marine angiosperms: evolutionary implications. *Glycobiology* **15**, 11–20 (2005).
5. Franssen, S. U. *et al.* Transcriptomic resilience to global warming in the seagrass *Zostera marina*, a marine foundation species. *Proc. Natl. Acad. Sci. USA* **108**, 19276–19281 (2011).
6. Mazzuca, S. *et al.* Establishing research strategies, methodologies and technologies to link genomics and proteomics to seagrass productivity, community metabolism, and ecosystem carbon fluxes. *Front. Plant Sci.* **4**, 1–19 (2013).
7. Duarte, C. M. *et al.* Will the oceans help feed humanity? *Bioscience* **59**, 967–976 (2009).
8. Costanza, R. *et al.* The value of the world's ecosystem services and natural capital. *Nature* **387**, 253–260 (1997).
9. Fourqurean, J. W. *et al.* Seagrass ecosystems as a globally significant carbon stock. *Nature Geosci.* **5**, 505–509 (2012).
10. Green, E. P. & Short, F. T. *World Atlas of Seagrasses* (University of California Press, Berkeley, CA, USA, 2003).
11. Wang, W. *et al.* The *Spirodela polyrhiza* genome reveals insights into its neotenous reduction fast growth and aquatic lifestyle. *Nature Commun.* **5**, 1–13 (2014).
12. Chavez Montes, R. A. *et al.* Sample sequencing of vascular plants demonstrates widespread conservation and divergence of microRNAs. *Nature Commun.* **5**, 1–15 (2014).
13. Vanneste, K., Maere, S. & Van de Peer, Y. Tangled up in two: a burst of genome duplications at the end of the Cretaceous and the consequences for plant evolution. *Philos. Trans. R. Soc. B Biol. Sci.* **369**, 20130353 (2014).
14. Nauheimer, L., Metzler, D. & Renner, S. S. Global history of the ancient monocot family Araceae inferred with models accounting for past continental positions and previous ranges based on fossils. *New Phytol.* **195**, 938–950 (2012).
15. Golicz, A. A. *et al.* Genome-wide survey of the seagrass *Zostera muelleri* suggests modification of the ethylene signalling network. *J. Exp. Bot.* (2015).
16. Kirk, J. T. O. in *Light and Photosynthesis in Aquatic Ecosystems* (Cambridge Univ. Press, 2011).
17. Touchette, B. W. Seagrass-salinity interactions: physiological mechanisms used by submersed marine angiosperms for a life at sea. *J. Exp. Mar. Biol. Ecol.* **350**, 194–215 (2007).
18. Popper, Z. A. *et al.* Evolution and diversity of plant cell walls: from algae to flowering plants. *Annu. Rev. Plant Biol.* **62**, 567–590 (2011).
19. Michel, G., Tonon, T., Scornet, D., Cock, J. M. & Kloareg, B. The cell wall polysaccharide metabolism of the brown alga *Ectocarpus siliculosus*: insights into the evolution of extracellular matrix polysaccharides in eukaryotes. *New Phytol.* **188**, 82–97 (2010).
20. Collen, J. *et al.* Genome structure and metabolic features in the red seaweed *Chondrus crispus* shed light on evolution of the Archaeplastida. *Proc. Natl. Acad. Sci. USA* **110**, 5247–5252 (2013).
21. Hanson, S. R., Best, M. D. & Wong, C. H. Sulfatases: structure, mechanism, biological activity, inhibition, and synthetic utility. *Angew. Chem. Int. Ed.* **43**, 5736–5763 (2004).
22. Larkum, A. W. D., Drew, E. A. & Ralph, P. J. in *Seagrasses: Biology, Ecology and Conservation* (eds Larkum, A. W. D., Orth, R. J. & Duarte, C. M.) 323–345 (Springer, Dordrecht, Netherlands, 2006).
23. De Cock, A. W. Flowering, pollinations and fruiting in *Zostera marina* L. *Aquat. Bot.* **9**, 201–220 (1980).
24. Furness, C. A. in *Early Events in Monocot Evolution* (eds Wilkin, P. & Mayo, S. J.) 1–22 (Cambridge Univ. Press, 2013).
25. Kuo, J. & den Hartog, C. in *Seagrasses: Biology, Ecology and Conservation* (eds Larkum, A. W. D., Orth, R. J. & Duarte, C. M.) 51–87 (Springer, 2006).
26. Orth, R. J. *et al.* A global crisis for seagrass ecosystems. *Bioscience* **56**, 987–996 (2006).
27. Waycott, M. *et al.* Accelerating loss of seagrasses across the globe threatens coastal ecosystems. *Proc. Natl. Acad. Sci. USA* **106**, 12377–12381 (2009).
28. Macreadie, P. I., Schliep, M. T., Rasheed, M. A., Chartrand, K. M. & Ralph, P. J. Molecular indicators of chronic seagrass stress: a new era in the management of seagrass ecosystems? *Ecol. Indic.* **38**, 279–281 (2014).
29. Vanneste, K., Baele, G., Maere, S. & Van de Peer, Y. Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous–Paleogene boundary. *Genome Res.* **24**, 1334–1347 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements Genome sequencing, assembly and automated annotation were conducted by the US Department of Energy (DOE) Joint Genome Institute, Walnut Creek California, USA and supported by the Office of Science of the US DOE, Community Sequencing Program award (2009) contract No. DE-AC02-05CH11231 to J.L.O. Further bioinformatics and annotation was supported in part by the Ghent University Multidisciplinary Research Partnership 'Bioinformatics: From Nucleotides to Networks' to Y.V.d.P. Y.V.d.P also acknowledges support from the European Union Seventh Framework Programme (FP7/2007–2013) under European Research Council Advanced Grant Agreement 322739–DOUBLE-UP. RNA-seq (Finnish clone genotype) was funded by the Marine Benthic Ecology and Evolution (MarBEE) group, within the former Centre for Ecological and Evolutionary Studies (now Groningen Institute for Evolutionary Life Sciences), University of Groningen to J.L.O.

RNA-seq (flower tissues) was funded by the Excellence Cluster, Future Ocean, Kiel to T.B.H.R. Participation of G.P. and E.D. was supported by the MIUR Italian Flagship project RITMARE (NRP 2011-2013). G.A.P. was supported by FCT-EXCL/AAG-GLO/0661/2012. We thank I. D. Gromicho, KAUST, for his artistry in the production of Extended Data Fig. 6. This work also benefited from discussions within the ESSEM COST action ES0906, "Seagrass productivity from genes to ecosystem management" (2009-2014), J.L.O., G.P. and T.B.H.R.; and the Linnaeus Centre for Marine Evolutionary Biology (CEMEB)-Tjärnö, Gothenburg University, J.L.O. and M.T. J.L.O. especially thanks K. Johannesson (CEMEB-Tjärnö), C. Boyen (SBR-Roscoff), R. Reinhardt (MPI-Cologne) and E. Serrão (CCMAR-Faro) for their ongoing encouragement, and the more than 70 colleagues who submitted letters of support for the original proposal to the JGI-Community Sequencing Program.

Author Contributions J.L.O., T.B.H.R., G.P. and Y.V.d.P. are the lead investigators and contributed equally to the work. J.S., J.W.J., J.G., Y.V.d.P., B.V. and Y.-C.L. coordinated the bioinformatics activities surrounding assembly, quality control, set-up and maintenance of *Z. marina* on the ORCAE site and deposition of the *Z. marina* genome resource. T.B.H.R. and T.B. generated and analysed RNA-seq libraries from flowers, rhizome, roots. J.L.O., Y.-C.L. and A.J. generated and analysed RNA-seq libraries from the genome genotype and temperature stress experiments. C.B., W.T.S. and J.L.O. contributed to biological sample collection, preparation and quality control prior to DNA extraction. A.M. performed the HMW DNA extraction and quality control from the genome genotype/clone. M.A., J.G., H.T. and M.C. contributed to WGS libraries and sequencing (fosmid)-cloning and quality control. J.G. coordinated the sequencing of FES, quality control projects. Analysis of architectural features of the genome and annotation of specific gene families, including the written contributions to the main paper and Supplementary Information sections, were performed by the following co-authors: J.W.J., the chromosome assembly analysis; B.V. and Y.-C.L., gene family clustering and comparative phylogenomics; A.R.K. and E.B.B., Pfam domains; E.D.P. and P.J.G., miRNA; R.L., K.V. and Y.V.d.P., whole-genome duplication; F.M., Y.-C.L. and Y.V.d.P., transposable elements; B.V., co-linearity and synteny comparisons; M.T., organellar genomes; P.R., stomata gene family; G.M., cell wall polysaccharides and sulfotransferases; T.T., fatty

acid metabolism and its relationship to cell walls and ion homeostasis; P.R., volatiles (ethylene, terpenes); P.R., J.B. and T.B.H.R., metallothioneins; P.R., G.A.P. and C.L., osmoregulation/ion homeostasis/stress-related genes; S.D. and E.D., photosynthetic/light-sensing genes; G.M., CAZymes; T.B., T.B.H.R. and P.R., plant defence-related; T.B. assembly and analysis of MADS box genes (flowering); P.R., Y.V.d.P. and Y.-C.L., pollen-related and self-incompatibility genes; F.M., SLR-1 gene and core eukaryotic genes analysis (CEGMA). J.L.O., Y.V.d.P., T.B.H.R., C.M.D., Y.-C.L. and P.R. wrote and edited the main manuscript (including the Methods and Extended Data), and organized and further edited the individual contributions (as listed above) for the Supplementary Information sections. J.L.O. and Y.V.d.P. provided the overall evolutionary context and T.B.H.R., G.P. and C.M.D. provided the ecological and societal context. All authors read and commented on the manuscript.

Author Information Raw reads, the assembled genome sequence and annotation are accessible from NCBI under BioProject number PRJNA41721 with GenBank accession number LFYR00000000. The accession number for the *Zostera marina* Finnish Clone is BioSample SAMN00991190. Fosmid end sequence: GSS KG963492-KG999999; K0000001-K0144970, whole-genome shotgun data: SRA020075 and RNA-seq: GEO GSE67579. Further information on the *Zostera marina* project is available via the Online Resource for Community Annotation Eukaryotes (ORCA) at <http://bioinformatics.psb.ugent.be/orcae/>. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.L.O. (j.l.olsen@rug.nl) or Y.V.D.P (yves.vandepeer@psb.vib-ugent.be).



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported licence. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons licence, users will need to obtain permission from the licence holder to reproduce the material. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/>.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Plant material and DNA preparation. A single genotype/clone of *Zostera marina* (referred to as the 'Finnish clone') was harvested on 26 August 2010 at 2 m depth at Färö Island (latitude 59° 55.234' N longitude 21° 47.766' E) located in the northern Baltic Sea, Finland. Plant material was transported to the lab in seawater, cleaned and further processed. Care was taken to use leaf-meristem tissue harvested from the inner layer of basal shoots to minimize bacterial/diatom contamination. Tissues were immediately frozen in LN₂ and stored at -80 °C for later DNA and RNA extraction. Monoclonality was verified by genotyping 40 ramets of the mega-clone with six highly polymorphic, microsatellite loci³⁰. There was no evidence for polyploidy^{25,31,32} (*Z. marina* is 2n = 12) or somatic mutations³³ as assessed by multiple peaks in the microsatellite chromatograms. Tissue was subsequently sent on dry ice to Amplicon Express for HMW DNA extraction using a CTAB isolation method modified by R. Meilan (unpublished) but available from him (rmeilan@purdue.edu), based on the original method³⁴. Following QC according to JGI guidelines, the DNA was shipped to JGI for library and sequencing preparation.

Genome sequencing and assembly. One 35-Kb, fosmid library was generated for end sequencing. The fosmid ends were sequenced with standard Sanger sequencing protocols at the HudsonAlpha Institute for a total of 194,303 Sanger reads (0.29× coverage). Illumina libraries (two fragment libraries (6.62 Gb), one 2-Kb JGI mate-pair library (3.57 Gb), one 4-Kb JGI mate-pair library (3.41 Gb) and two 8-Kb JGI mate-pair libraries (11.94 Gb)) were sequenced with Illumina MiSeq/HiSeq genetic analysers at the Department of Energy's Joint Genome Institute (JGI), using standard protocols. A total of 25.55 Gb of Illumina and 0.14 Gb of Sanger sequence was obtained representing 47.7× genomic coverage. Prior to assembly, all reads were screened against mitochondria, chloroplast, and Illumina controls. Reads composed of > 95% simple sequence repeats were removed. For the Illumina, paired-end libraries (2 × 250), reads < 75 bp were discarded, for the 2 × 150 libraries, reads < 50 bp were discarded after trimming for adaptor and quality ($q < 20$). An additional deduplication step was performed on the mate pairs that identified and retained only one copy of each PCR duplicate. A total of 212,101,273 reads (Supplementary Table 2.1) was assembled using our modified version of Arachne v. 20071016 (ref. 35). Subsequent directed Arachne modules were applied to collapse adjacent heterozygous contigs. The entire assembly was then run through another Arachne process starting at Stage 6 Rebuilder. This produced 15,747 scaffold sequences (30,723 contigs), with a scaffold L50 of 409.5 Kb, 613 scaffolds larger than 100 Kb, and a total genome size of 237.5 Mb (Supplementary Table 2.2).

Scaffolds were screened against bacterial proteins, organelle sequences, GenBank NR (nr_prot) and RefSeq protein databases, and removed if found to be a contaminant. Scaffolds consisting of prokaryotes, chloroplast, mitochondria and unanchored rDNA were removed. We also assembled the chloroplast and partial mitochondrial genomes (Supplementary Notes 2.2 and 2.3, Supplementary Fig. 2.1). Additionally, short (<1 Kb) scaffolds or scaffolds containing highly repetitive sequence (> 95% 24-mers found more than four times in large scaffolds) or alternative haplotypes were also removed. Following repeat analysis and gene prediction, all scaffolds were subjected to a filtering process (based on NCBI nr_prot + NCBI taxonomy database) to eliminate remaining bacterial (and other) contaminants (Supplementary Table 2.3).

Assembly validation was performed using a set of 12 fully sequenced fosmid clones. In 4 of the 12 fosmid clones, full-length alignments were not found due to fragmentation in the region of the fosmid clone. In five of the remaining eight fosmid clones, the alignments were of high quality (< 0.05% bp error). The overall base pair error rate (including marked gap bases) in the fosmid clones that aligned to full length was 0.28% (714 discrepant base pairs out of 253,332 bp). Supplementary Table 2.4 shows the individual fosmid clones and their contribution to the overall error rate. Note that two fosmid clones (16248, 16249) contributed nearly 81% of the discrepant bases. This probably occurred in polymorphic regions of the genome where the haplotype in the fosmid did not match the haplotype in the reference. There are several indels of various sizes in the clone and assembly, typical of a region of degraded transposons. Further quality analysis indicated that 90% of the set of eukaryotic core genes (CEGMA) were present and 98% were partially represented, suggesting near completeness of the euchromatin component.

Annotation of repetitive sequences. Two complementary approaches were used to identify repetitive DNA sequences in the *Z. marina* genome. With respect to masking repeats before gene prediction analysis, a *de novo* repeat identification was carried out with RepeatModeler (v. open-1.0.7; <http://www.RepeatMasker.org>)³⁶ to identify repeat boundaries and build consensus models from which potential over represented, non-transposable element, protein-coding genes were removed. RepeatMasker (v. open-4.0.0, WUblast) was used in combination with

this custom repeat library to mask the assembly and prepare it for gene prediction with EuGene.

Furthermore, in order to perform a qualitative and quantitative analysis of repeats with greater resolution³⁷ the genome assembly was processed for *de novo* repeat detection using the TEedenovo pipeline from the REPET package v. 2.2 (ref. 38); parameters were set to consider repeats with at least five copies. The consensus sequences generated by TEedenovo were then used as probes for whole genome annotation by the TEannot³⁹ pipeline from the REPET package v. 2.2. The consensus repeat sequences were classified using PasteC⁴⁰. Comparing the genomic positions of transposable elements (TE) to those of exons from the set of predicted genes enabled us to identify that 909 gene predictions most likely represent TEs and these were filtered from the gene set. The REPET package v. 2.2 was also used to annotate repetitive elements in the *Spirodela polyrhiza* genome assembly with the same parameters as for *Z. marina*. See Supplementary Fig. 3.1.

Transcriptome library preparation, sequencing and assembly. Leaf, root and flower tissues were separately frozen in liquid nitrogen immediately following harvest from either ambient (field collected) or experimental (mesocosm) conditions (Supplementary Note 3.2). Overall, we obtained between nine and 20 million high-quality reads from each of the flower-leaf-root replicate libraries; and for the Finnish clone library, 148.5 million high quality reads were retrieved (Supplementary Table 3.3).

The *de novo* assembly protocol was adapted from ref. 41. We pooled replicates of each tissue together except for the two leaf tissue libraries, which were kept separate (Supplementary Table 3.4) and performed *de novo* transcriptome assembly for each tissue using Trinity⁴¹ (v. 2014-07-17) with digital normalization option ON to normalize input read coverage. Frame shift errors and insertion/deletion errors in the assembled transcripts were corrected by FrameDP⁴². Because a *de novo* assembly still generates many spurious transcripts, we used the transcript expression value to remove low quality contigs. We used the RSEM pipeline⁴³ to obtain the contig expression values and removed contigs with FPKM (fragments per kilobase of transcript per million fragments mapped) value < 1 and IsoPct (percentage of expression for a given transcript compared with all expression from that Trinity component) < 1. In total, we obtained between 39,000 and 53,000 assembled contigs from each library, and 52,000 contigs from the Finnish clone library (Supplementary Table 3.4). Prior to mapping the genome sequence and the predicted genes, we used the CD-HIT⁴⁴ program (v. 4.6.1) to collapse redundant contigs, which resulted in 79,134 low redundant transcript contigs.

Differential gene expression analysis. High-quality RNA-seq reads were mapped to the genome assembly v.2.1 by TopHat⁴⁵. Differential gene expression analysis was performed by the Cufflink pipeline⁴⁵ based on the *Z. marina* v.2.1 gene models by converting the number of aligned reads into FPKM values. Genes with significant expression difference ($\log_2 > 2$) were selected for further investigation by GOSTats⁴⁶ to perform Gene Ontology (GO) term enrichment analysis with $P \leq 0.05$ (Supplementary Note 3.3, Supplementary Table 3.5).

MicroRNA analysis. Genomic precursors of known miRNAs were mapped on the *Z. marina* genome following the procedure described in ref. 47 for the maize genome. miRNA entries from the miRBase database (release 21, 2014) were aligned to the chromosomes of the *Z. marina* genome. Up to three mismatches were allowed in the alignment, using SeqMap⁴⁸. In parallel, novel potential DCL1/AGO1-dependent miRNAs were enriched by selecting 5'-U 20–22 nt small RNAs from three different sequenced libraries from *Z. marina* described in ref. 12. A subset of these small RNAs with abundance ≥ 10 TPM (transcripts per million) was retained and aligned to the genome with no mismatches. From every locus, we extracted two ~200-nt regions surrounding each aligned miRNA or candidate (from -30 to +160 and from -160 to +30 nucleotides relative to the putative miRNA start or end coordinate, respectively). Minimum energy RNA secondary structures were predicted for each region using the RNAfold program of the Vienna RNA 1.8.5 package (<http://www.tbi.univie.ac.at/~ivo/RNA/>) using default settings.

In addition, small RNAs from the three sequenced libraries were mapped on these regions, allowing no mismatches, in order to pre-select putative miRNA loci that showed evidence of expression in the three plant tissues analysed. We evaluated RNA structure and small RNA alignment in all the regions based on: (1) dominance of plus-stranded small RNAs; (2) position of the most abundant small RNAs relative to the predicted miRNA coordinates; (3) prevalence of 20–22 nt small RNAs in the predicted miRNA locus; (4) position of the putative miRNA with the stem-loop structure; and (5) absence of oversize (≥ 3 nt) bulges in the miRNA/miRNA* alignment. After reduction of overlapping loci to a non-redundant set and removal of stem-loop structures with the wrong orientation compared to miRNAs registered in miRBase, we manually inspected the remaining loci to further evaluate them according to the miRNA annotation criteria proposed by ref. 49. Stringency was relaxed when small RNA expression data strongly indicated the presence of miRNA loci that did not meet the whole set of criteria. Novel miRNA precursors overlapping with TEs or other repetitive elements were filtered out.

Potential miRNA targets were identified *in silico* using the generic small RNA-transcriptome aligner GSTAr from the CleaveLand package (v. 4)⁵⁰. Predicted targets were accepted with an Allen score <4 or a MFE (minimum free energy) ratio ≥ 7.5 . (Supplementary Note 3.4).

Gene prediction. Training of the gene prediction programs started with the collection of high quality ESTs. EST information was used, for example, to train the splice predictor SpliceMachine⁵¹. Detection of conserved splice sites was further investigated by RNA-seq splice junctions (count > 10) to construct a WAM model in EuGene (v. 4.1)⁵². Coding-potential was modelled with an interpolated Markov Model (IMM) constructed from the BLASTX alignments of proteins from the PLAZA v. 2.5 database⁵³. An additional protein 'monocot' Markov Model was built based on the protein sequences from *Brachypodium*, maize and sorghum. Starting from EST and protein alignments, a set of 215 gene models was manually constructed and curated using the genome browser GenomeView⁵⁴. The 215 models were then used as a training set for EuGene in order to optimize the different splice site and coding-potential models, as well as the weights for the extrinsic EST and homology evidence. An overall fitness score of 80.1% was achieved, which is high enough to obtain reliable results without overfitting. GeneMark⁵⁵ and Augustus⁵⁶ were separately trained (using the same input data as EuGene) and their predictions were integrated with EuGene using a custom script to evaluate the best gene structure at each locus. All gene models were automatically screened to highlight possible erroneous structures (for example, in-frame stop codons, deviating splice junctions) and manually curated. Transfer-RNA gene models were predicted by tRNAscan-SE (v. 1.31)⁵⁷ and their structures were verified with Infernal (v. 1.1rc1, rfam11 covariant model database)⁵⁸. For each gene, UTRs were assigned by identifying a set of ESTs and RNA-seq assemblies that uniquely overlapped with it. We subsequently selected the longest mapped transcript on either end of the predicted coding sequence and designated the section outside the coding sequence as the UTR. Finally, all genes were uploaded to the ORCAE platform (<http://bioinformatics.psb.ugent.be/orcae>)⁵⁹, enabling all members of the consortium to refine and curate the gene model and assign gene function. A list of protein domains, as well as the derived Gene Ontology (GO) terms and KEGG pathway identifiers were generated using an InterProScan (v. 5.2.45)⁶⁰ analysis and available in ORCAE. More specifically, gene functional descriptions were added either manually by consortium expert scientists or automatically through sequence homology searches. The automated method relies on the EC (Enzyme Commission) number reported by InterProScan to retrieve the enzyme name with BLASTP search against UniProtKB/Swiss-Prot⁶¹ to filter out hits that are below 60% identity and 70% query/hit coverage. Although such high stringency on per cent identity and sequence coverage reduced the available number of functional descriptions, it reduced the false-positive prediction rate, as desired here.

Construction of age distributions and WGD analyses. K_S -based age distributions were constructed as previously described⁶². In brief, the K_S values between genes were obtained through maximum likelihood estimation using the CODEML program⁶³ of the PAML package (v. 4.4c)⁶⁴. Gene families for which K_S estimates between members did not exceed a value of 5 were subdivided into subfamilies. For each duplicated gene in the resulting phylogenetic gene tree, obtained by PhyML⁶⁵, all K_S estimates between the two child clades were added to the K_S distribution with a weight $1/m$ (where m is the number of K_S estimates for a duplication event), so that the weights of all K_S estimates for a single duplication event summed to one. Mixture modelling was used to confirm a WGD signature in the K_S distribution (Fig. 2 and Supplementary Fig. 4.1), for which all duplicates with K_S values ≤ 0.1 were excluded to avoid the incorporation of allelic and/or splice variants, while all duplicates with K_S values > 2.0 were removed because K_S saturation and stochasticity can mislead mixture modelling above this range⁶². For further details see Supplementary Note 4.1.

Absolute dating of the identified WGD event was performed as described previously^{13,29}. In brief, paralogous gene pairs located in duplicated segments (anchors) and duplicated pairs lying under the WGD peak (peak-based duplicates) were collected for phylogenetic dating. Anchors, assumed to be corresponding to the most recent WGD, were detected using i-ADHoRe 3.0 (refs 66,67). Only a low number of duplicated segments and hence anchors could be identified, most likely because of the fragmented assembly of *Z. marina*. However, the identified anchors did confirm the presence of a broad WGD peak between a K_S of 0.8 and 1.6 (data not shown). For each WGD paralogous pair, an orthogroup was created that included the two paralogues plus several orthologues from other plant species as identified by InParanoid (v. 4.1)⁶⁸ using a broad taxonomic sampling: one representative orthologue from the order Cucurbitales, two from the Rosales, two from the Fabales, two from the Malpighiales, two from the Brassicales, one from the Malvales, one from the Solanales, two from the Poales, one orthologue from *Musa acuminata*⁶⁹ (Zingiberales), and one orthologue from *Spirodela polyrhiza*¹¹ (Alismatales). In total, about 180 orthogroups from anchor pair duplicates and peak-based duplicates were collected. The node joining the two *Z. marina* WGD paralogues was then dated using the BEAST v. 1.7 package⁷⁰ under an uncorrelated

relaxed clock model and a LG+G (four rate categories) evolutionary model. A starting tree with branch lengths satisfying all fossil prior constraints was created according to the consensus APGIII phylogeny⁷¹. Fossil calibrations were implemented using log-normal calibration priors on the following nodes: the node uniting the Malvidae based on the fossil *Dressiantha bicarpellata*⁷² with prior offset = 82.8, mean = 3.8528, and s.d. = 0.5 (ref. 73), the node uniting the Fabidae based on the fossil *Paleoclusia chevalieri*⁷⁴ with prior offset = 82.8, mean = 3.9314, and s.d. = 0.5 (ref. 75), the node uniting the Alismatales (including *Z. marina* and *Spirodela polyrhiza*) with the other monocots based on the oldest fossil monocot pollen, *Liliacidites*^{76,77} from the Trent's Reach locality, with prior offset = 125, mean = 2.0418, and s.d. = 0.5 (refs 14,78) and the root with prior offset = 124, mean = 4.0786, and s.d. = 0.5 (ref. 79). The offsets of these calibrations represent hard minimum boundaries, while their means represent locations for their respective peak mass probabilities in accordance with some of the most recent and taxonomically complete dating studies available for these specific clades^{14,80}. A run without data was performed to ensure proper placement of the marginal calibration prior distributions⁸¹. The Markov chain Monte Carlo (MCMC) for each orthogroup was run for 10^6 generations, sampling every 1,000 generations resulting in a sample size of 10^4 . The resulting trace files of all orthogroups were evaluated manually using Tracer v. 1.5⁷⁰ with a burn-in of 1,000 samples to ensure proper convergence (minimum ESS for all statistics at least 200). In total, 169 orthogroups were accepted and all age estimates for the node uniting the WGD paralogous pairs were then grouped into one absolute age distribution (Fig. 2, too few anchors were available to evaluate them separately from the peak-based duplicates), for which kernel density estimation (KDE) and a bootstrapping procedure were used to find the peak consensus WGD age estimate and its 90% confidence interval boundaries, respectively.

Intra- and inter-genomic co-linearity was investigated (Supplementary Tables 4.2 and 4.3) using MCScanX⁸² based on a BLASTP search of all genomic protein coding genes with an E-value cut-off of e^{-10} . Only one large duplicated segment was detected, which was most likely due to the fragmented assembly of *Z. marina*; only 27 scaffolds had a size larger than 1 Mb, accounting for only 23.4% of all protein-coding genes. We therefore additionally used i-ADHoRe (v. 3.0)⁶⁶ to investigate genomic co-linearity by including all possible scaffolds.

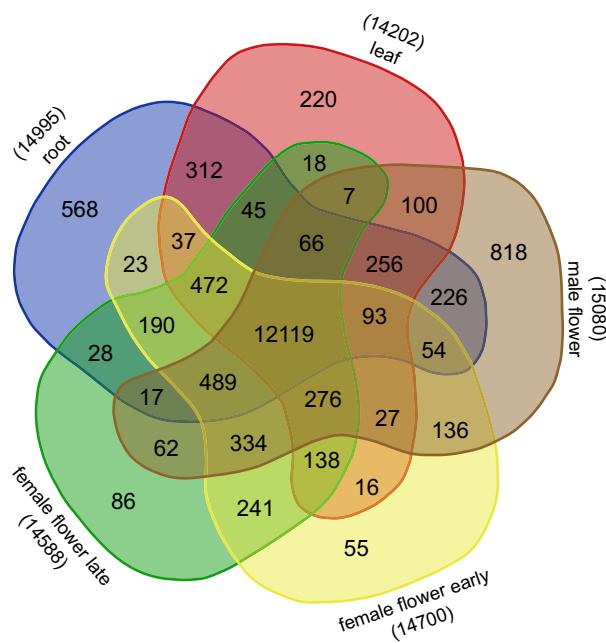
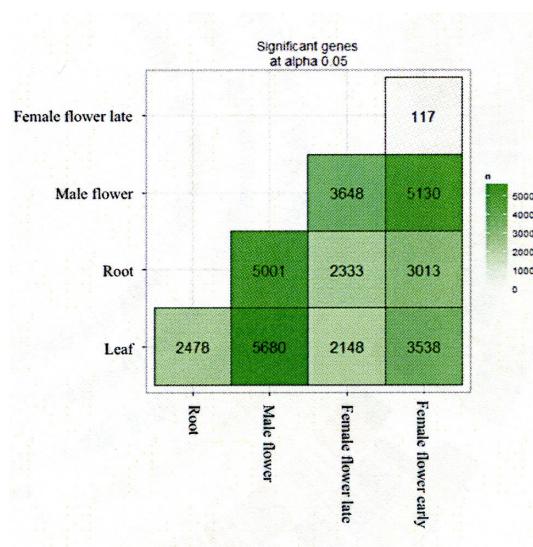
Gene family comparisons. Protein sets were collected for 14 species: *Z. marina* (ORCAE v. 2.1), *Arabidopsis thaliana* (TAIR10), *Thellungiella parvula* (<http://thellungiella.org>), *Populus trichocarpa* (Phytozome v. 9.0), *Vitis vinifera* (Phytozome v. 9.0), *Amborella trichopoda* (<http://amborella.huck.psu.edu>), *Oryza sativa* japonica (Phytozome v. 9.0), *Zea mays* (Phytozome v. 9.0), *Brachypodium distachyon* (Phytozome v. 9.0), *Spirodela polyrhiza* (<http://mocklerlab.org>), *Selaginella moellendorffii* (Phytozome v. 9.0), *Physcomitrella patens* (Phytozome v. 9.0), *Chlamydomonas reinhardtii* (Phytozome v. 9.0), and *Ostreococcus lucimarinus* (ORCAE v. 6/3/2013). These species were selected in order to provide a phylogenetic representation traversing green algae, basal plants, monocots, and dicots. Following an 'all-vresus-all' TimeLogic Decypher Tera-BLASTP (Active Motif Inc.; e-value threshold $1 \times e^{-3}$, max hits 500) comparison, OrthoMCL (v. 2.0; mcl inflation factor 3.0)⁸³ was used to delineate gene families. Confidence in establishing gene losses in *Zostera* was enhanced by using a combination of reciprocal blast, TblastN, re-annotation of *Spirodela* (and other monocot genes), and careful phylogenetic analysis. OrthoMCL results and related protein resources are available in the ORCAE download section.

To further understand gene family expansion or contraction in *Z. marina* in comparison with other sequenced genomes, gene family sizes were calculated for all gene families (excluding orphans and species-specific families) (Supplementary Note 4.2). The number of genes per species for each family was transformed into a matrix of z-scores in order to centre and normalize the data. The first 100 families with the largest gene family size in *Z. marina* were selected. The z-score profile was hierarchically clustered (complete linkage clustering) using Pearson correlation as a distance measure. The functional annotation of each family was predicted based on sequence similarity to entries in the InterProScan and Pfam protein domain database where more than 30% of proteins in the family share the same protein domain. The phylogenetic profile and phylogenetic tree topology provided at PLAZA⁸⁴ were used to reconstruct the most parsimonious series of gene gain and loss events. The Dollop program from the PHYLIP package⁸⁵ was used to determine the minimum gene set at ancestral nodes of the phylogenetic tree. The Dollop program is based on the Dollo parsimony principle, which assumes that novel gene families arise exactly once during evolution but can be lost independently in different phylogenetic lineages.

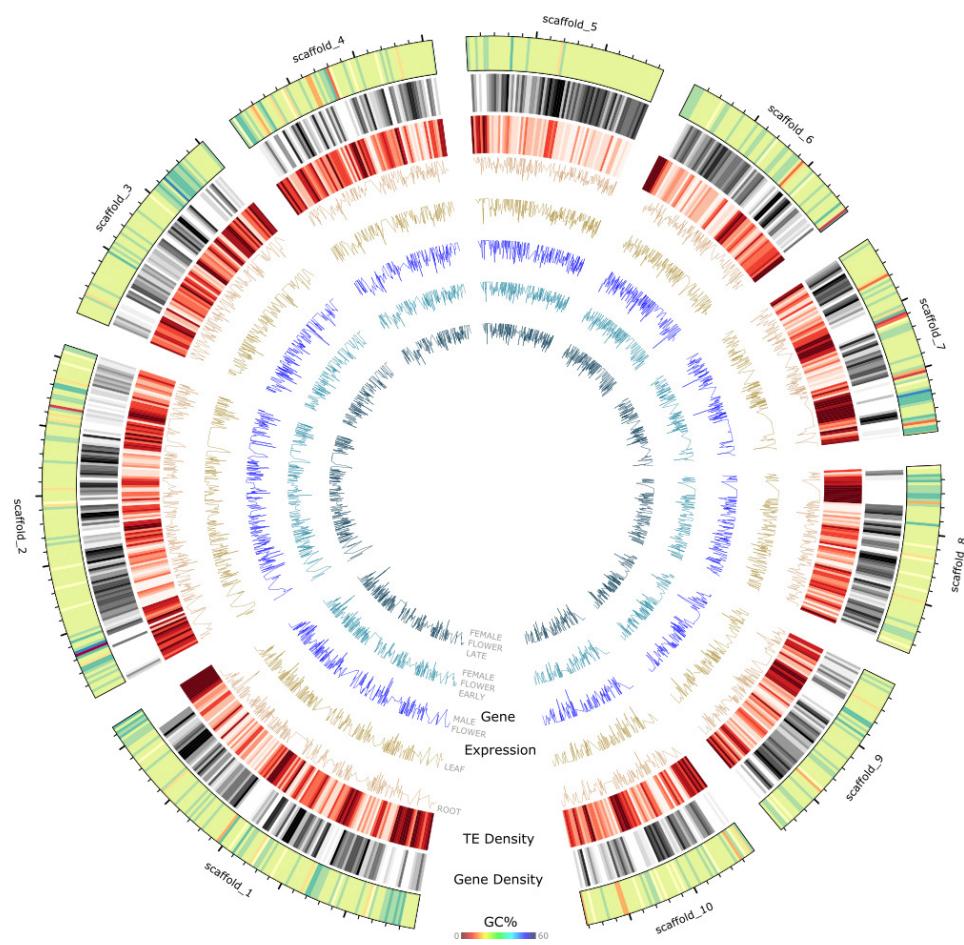
Search for presence/absence of orthologues for specific genes and families. A dedicated search for orthologues/homologues was performed for genes and proteins involved in stomata differentiation (Supplementary Note 5.1), volatile biosynthesis and sensing with focus on ethylene and terpenes (Supplementary Note 6.1), as well as genes involved in male flower specification and pollen

differentiation (Supplementary Note 11.1). To this end, queries were chosen from documented genes involved in these pathways (usually from *Arabidopsis* but occasionally from *Oryza*, *Zea* and tomato). Next, the search for homologues in *Zostera marina*, *Spirodela polyrhiza*, *Oryza sativa japonica* and *Arabidopsis thaliana* (when not used as a query) was performed using BLASTP. To avoid missing or poorly annotated genes a TBLASTN search was conducted using the above queries against the *Zostera marina* and *Spirodela polyrhiza* genomes. Putative orthologues were identified based on reciprocal BLASTP searches with *Arabidopsis* (or the other queries). Owing to species-specific duplications, this may produce some paralogous genes to appear orthologous to the query, or vice versa (see Extended Data Tables 1–3). To further confirm correct orthology assignments, phylogenetic trees were built using a broader sampling of protein sequences from both the query species and the three target species. Ambiguously aligned sequences (especially due to indels) were checked manually and corrected or removed.

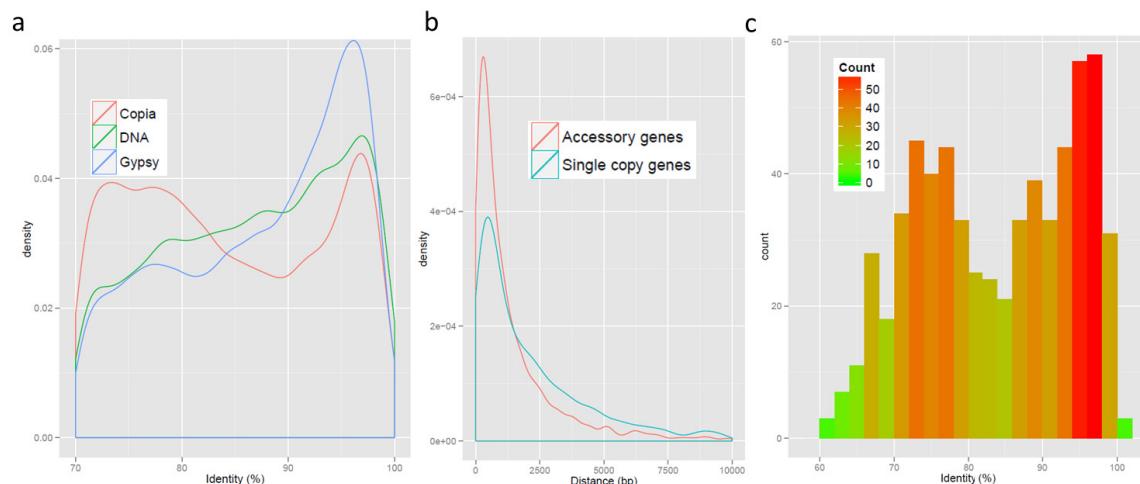
30. Olsen, J. L. *et al.* Eelgrass *Zostera marina* populations in northern Norwegian fjords are genetically isolated and diverse. *Mar. Ecol. Prog. Ser.* **486**, 121–132 (2013).
31. den Hartog, C., Hennen, J., Noten, T. M. P. A. & Van Wijk, R. J. Chromosome numbers of the European seagrasses. *Plant Syst. Evol.* **156**, 55–59 (1987).
32. Kuo, J. Chromosome numbers of the Australian Zosteraceae. *Plant Syst. Evol.* **226**, 155–163 (2001).
33. Reusch, T. B. H. & Bostrom, C. Widespread genetic mosaicism in the marine angiosperm *Zostera marina* is correlated with clonal reproduction. *Evol. Ecol.* **25**, 899–913 (2010).
34. Doyle, J. J. & Doyle, J. L. Isolation of plant DNA from fresh tissue. *Focus* **12**, 13–15 (1990).
35. Jaffe, D. B. *et al.* Whole-genome sequence assembly for mammalian genomes: Arachne 2. *Genome Res.* **13**, 91–96 (2003).
36. Smit, A. & Hubley, R. in *RepeatModeler Open-1.0* (Repeat Masker Website, <http://www.repeatmasker.org/> 2010).
37. Maumus, F. & Quesneville, H. Deep investigation of *Arabidopsis thaliana* junk DNA reveals a continuum between repetitive elements and genomic dark matter. *PLoS ONE* **9**, e94101 (2014).
38. Flutre, T., Duprat, E., Feuillet, C. & Quesneville, H. Considering transposable element diversification in *de novo* annotation approaches. *PLoS ONE* **6**, e16526 (2011).
39. Quesneville, H. *et al.* Combined evidence annotation of transposable elements in genome sequences. *PLOS Comput. Biol.* **1**, e22 (2005).
40. Hoede, C. *et al.* PASTEC: an automatic transposable element classification tool. *PLoS ONE* **9**, e91929 (2014).
41. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnol.* **29**, 644–652 (2011).
42. Gouzy, J., Carrere, S. & Schiex, T. FrameDP: sensitive peptide detection on noisy matured sequences. *Bioinformatics* **25**, 670–671 (2009).
43. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
44. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
45. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
46. Falcon, S. & Gentleman, R. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**, 257–258 (2007).
47. Zhang, L. *et al.* A genome-wide characterization of microRNA genes in maize. *PLoS Genet.* **5**, e1000716 (2009).
48. Jiang, H. & Wong, W. H. SeqMap: mapping massive amount of oligonucleotides to the genome. *Bioinformatics* **24**, 2395–2396 (2008).
49. Meyers, B. C. *et al.* Criteria for annotation of plant microRNAs. *Plant Cell* **20**, 3186–3190 (2008).
50. Addo-Quaye, C., Miller, W. & Axtell, M. J. CleaveLand: a pipeline for using degradome data to find cleaved small RNA targets. *Bioinformatics* **25**, 130–131 (2009).
51. Degroeve, S., Saeyns, Y., De Baets, B., Rouze, P. & Van de Peer, Y. SpliceMachine: predicting splice sites from high-dimensional local context representations. *Bioinformatics* **21**, 1332–1338 (2005).
52. Foissac, S. *et al.* Genome annotation in plants and fungi: EuGene as a model platform. *Curr. Bioinformatics* **3**, 87–97 (2008).
53. Van Bel, M. *et al.* Dissecting plant genomes with the PLAZA comparative genomics platform. *Plant Physiol.* **158**, 590–600 (2012).
54. Abeel, T., Van Parys, T., Saeyns, Y., Galagan, J. & Van, P. GenomeView: a next-generation genome browser. *Nucleic Acids Res.* **40**, e12 (2012).
55. Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y. O. & Borodovsky, M. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res.* **18**, 1979–1990 (2008).
56. Stanke, M., Tzvetkova, A. & Morgenstern, B. AUGUSTUS at EGASP: using EST, protein and genomic alignments for improved gene prediction in the human genome. *Genome Biol.* **7**, S11 (2006).
57. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964 (1997).
58. Burge, S. W. *et al.* Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res.* **41**, D226–D232 (2013).
59. Sterck, L., Billiau, K., Abeel, T., Rouzé, P. & Van der Peer, Y. ORCAE: online resource for community annotation of eukaryotes. *Nature Methods* **9**, 1041 (2012).
60. Mitchell, A. *et al.* The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res.* **43**, D213–D221 (2015).
61. The UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res.* **43**, D204–D212 (2015).
62. Vanneste, K., Van de Peer, Y. & Maere, S. Inference of genome duplications from age distributions revisited. *Mol. Biol. Evol.* **30**, 177–190 (2013).
63. Goldman, N. & Yang, Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* **11**, 725–736 (1994).
64. Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
65. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
66. Proost, S. *et al.* i-ADHoRe 3.0-fast and sensitive detection of genomic homology in extremely large data sets. *Nucleic Acids Res.* **40**, e11 (2012).
67. Fostier, J. *et al.* A greedy, graph-based algorithm for the alignment of multiple homologous gene lists. *Bioinformatics* **27**, 749–756 (2011).
68. Ostlund, G. *et al.* InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. *Nucleic Acids Res.* **38**, D196–D203 (2010).
69. D'Hont, A. *et al.* The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* **488**, 213–217 (2012).
70. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. 3408070; Bayesian phylogenetics with BEAUTi and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969–1973 (2012).
71. The Angiosperm Phylogeny Group. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.* **161**, 105–121 (2009).
72. Gandolfo, M., Nixon, K. & Crepet, W. A new fossil flower from the Turonian of New Jersey: *Dressiantha bicarpellata* gen. et sp. nov. (Capparales). *Am. J. Bot.* **85**, 964–974 (1998).
73. Beilstein, M. A., Nagalingum, N. S., Clements, M. D., Manchester, S. R. & Mathews, S. Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **107**, 18724–18728 (2010).
74. Crepet, W. & Nixon, K. C. Fossil Clusiaceae from the late Cretaceous (Turonian) of New Jersey and implications regarding the history of bee pollination. *Am. J. Bot.* **85**, 1122–1133 (1998).
75. Xi, Z. *et al.* Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. *Proc. Natl. Acad. Sci. USA* **109**, 17519–17524 (2012).
76. Doyle, J. A., Endress, P. K. & Upchurch, G. R. Early Cretaceous monocots: a phylogenetic evaluation. *Acta Musei Nationalis Pragae, Series B. Historia Naturalis* **64**, 59–87 (2008).
77. Iles, W. J. D., Smith, S. Y., Gandolfo, M. A. & Graham, S. W. Monocot fossils suitable for molecular dating analyses. *Bot. J. Linn. Soc.* **178**, 346–374 (2015).
78. Janssen, T. & Bremer, K. The age of major monocot groups inferred from 800+ rbcL sequences. *Bot. J. Linn. Soc.* **146**, 385–398 (2004).
79. Smith, S. A., Beaulieu, J. M. & Donoghue, M. J. An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants. *Proc. Natl. Acad. Sci. USA* **107**, 5897–5902 (2010).
80. Clarke, J. T., Warnock, R. C. & Donoghue, P. C. Establishing a time-scale for plant evolution. *New Phytol.* **192**, 266–301 (2011).
81. Heled, J. & Drummond, A. J. Calibrated tree priors for relaxed phylogenetics and divergence time estimation. *Syst. Biol.* **61**, 138–149 (2012).
82. Wang, Y. *et al.* MCSanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49 (2012).
83. Li, L., Stoeckert, C. J. & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003).
84. Proost, S. *et al.* PLAZA 3.0: an access point for plant comparative genomics. *Nucleic Acids Res.* **43**, D974–D981 (2015).
85. Felsenstein, J. in *PHYLIP: Phylogenetic inference program*, version 3.6 (University of Washington, 2005).
86. Pillitteri, L. J. & Dong, J. Stomatal development in *Arabidopsis*. *Arabidopsis Book* **11**, e0162 (2013).
87. Lallemand, B., Erhardt, M., Heitz, T. & Legrand, M. Sporopollenin biosynthetic enzymes interact and constitute a metabolon localized to the endoplasmic reticulum of tapetum cells. *Plant Physiol.* **162**, 616–625 (2013).

a**b**

Extended Data Figure 1 | Number of genes expressed in five tissues of *Z. marina*. **a**, Venn diagram of genes with expression values (FPKM) higher than 1 are considered as expressed in the tissue. **b**, Pairwise differential gene expression analysis between tissues. The male flower shows the highest number of differentially expressed genes.

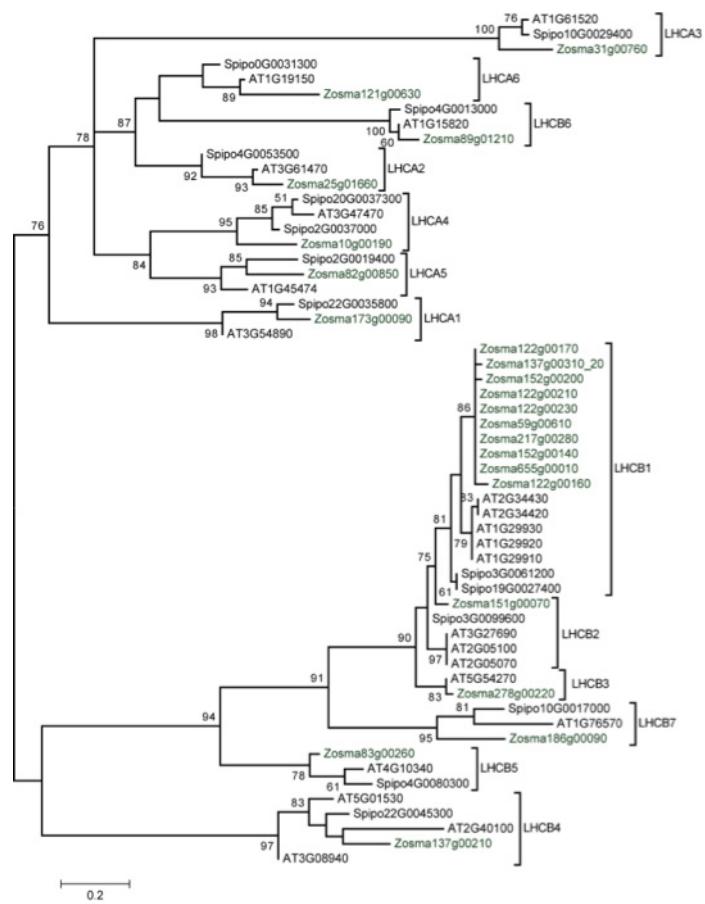


Extended Data Figure 2 | Circos plot of the ten largest scaffolds of *Z. marina*. Tracks from outside to inside. GC percentage, gene density, and transposable element (TE) density (density measured in 20-Kb sliding windows and gene expression profiles from five tissues (root, leaf, male flower, female flower early and female flower late) presented as \log_2 FPKM values.

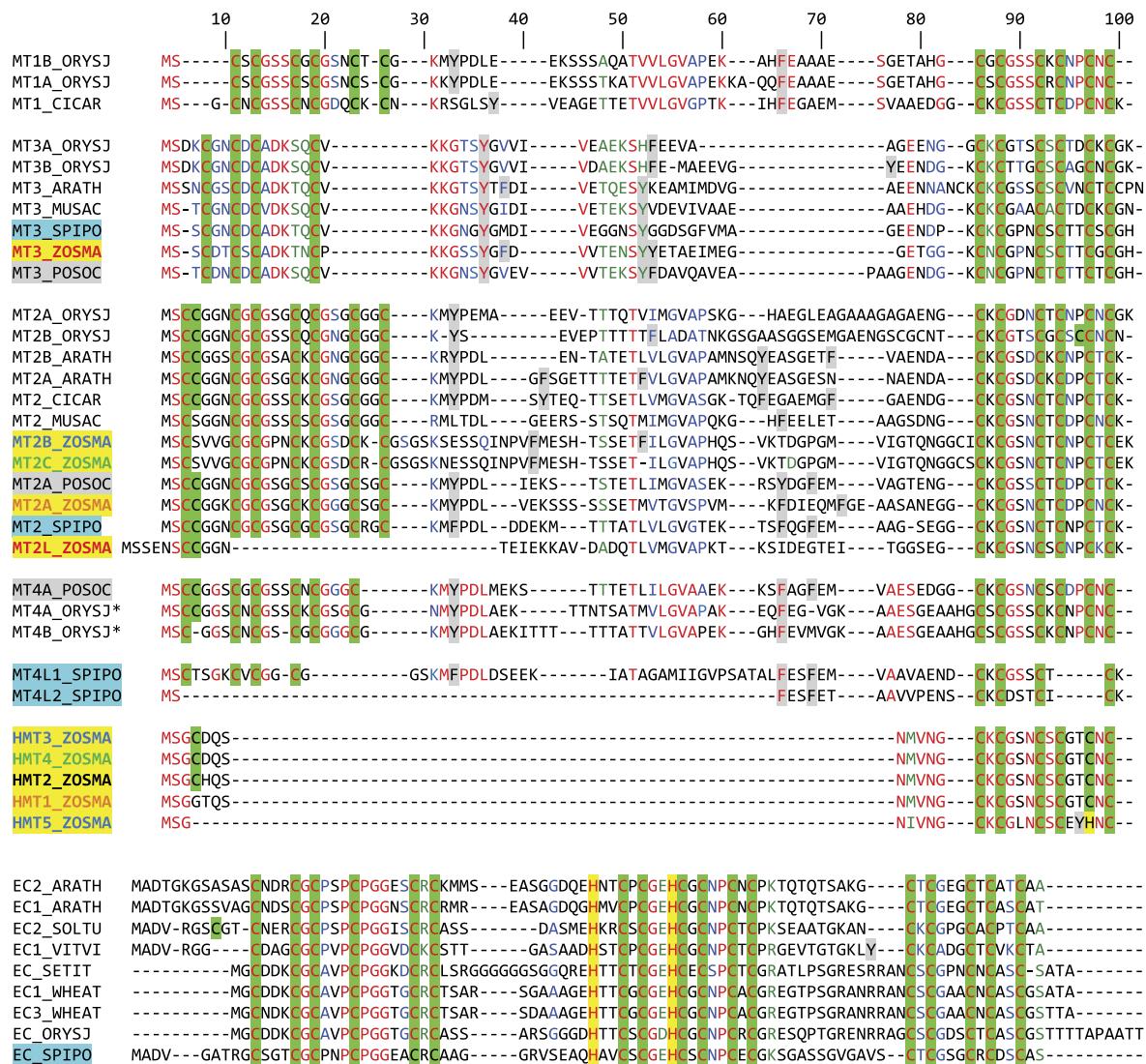


Extended Data Figure 3 | Potential impact of transposable elements (TEs) on *Z. marina* evolution. **a**, Frequency distribution of pairwise sequence identity values between copies of Copia- and Gypsy-type LTR retrotransposons and DNA transposons, and their cognate consensus sequences (younger repeats share higher sequence similarity). Two peaks are detectable for Copia-type elements. **b**, Distance to the closest TE for

the set of *Z. marina* single-copy genes and the set of *Z. marina* accessory genes. TE-proximal accessory genes are more frequent than TE-proximal single-copy genes. **c**, Frequency of pairwise sequence identity between accessory gene-proximal Ty3-Gypsy elements and their cognate consensus sequences. A number of high-identity copies (that is, putatively young duplicate genes) is observed.

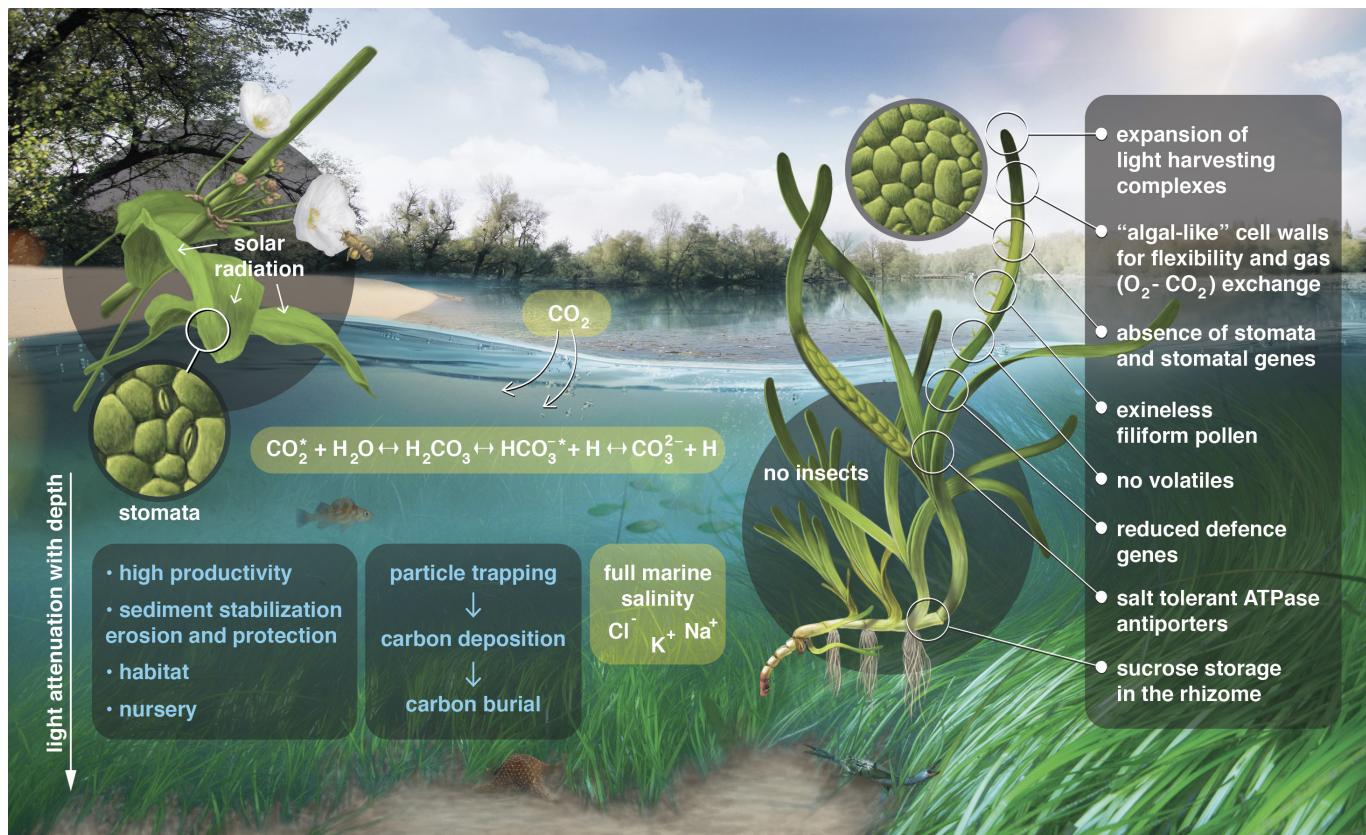


Extended Data Figure 4 | Unrooted maximum likelihood tree of genes encoding light-harvesting complex A (LHCA) and LHC B proteins of *Z. marina*, *Spirodesla polyrhiza* and *Arabidopsis thaliana*. The analysis was carried out on protein sequences using PhyML 3 with LG substitution model and 100 bootstrap replicates. Supplementary Note 7.1, Supplementary Table 7.3.



Extended Data Figure 5 | Alignment of metallothionein (MT) and half-m metallothionein (HMT) genes in *Z. marina* as compared with other plants. Alignments were performed in ClustalW on the Lyon PBIL web server and edited manually. The upper alignments are for type 1–3 MTs and HMTs; the lower alignment is for type 4 EcMTs where there is no *Zostera* homologue. Conserved residues are shown in red and residues in the same amino acid group in blue. Cys and His residues, putatively involved in binding metals, are highlighted in green and yellow,

respectively. Aromatic amino acids absent in canonical animal MTs are highlighted in grey. MTs and MT-like proteins were obtained from: *Arabidopsis thaliana* (ARATH), Japanese rice (ORYSJ), *Cicer arietinum* (CICAR), banana (MUSAC), wheat (WHEAT), potato (SOLTU), *Setaria Italica* (SETIT), *Vitis vinifera* (VITVI) and the alismatids: *Posidonia oceanica* (POSOC) highlighted in grey, *Spirodela polyrhiza* (SPIPO) highlighted in blue, and *Zostera marina* (ZOSMA) highlighted in yellow. See Supplementary Note 8.2.



Extended Data Figure 6 | Conceptual summary of physiological and structural adaptations made by *Z. marina* in its return to the sea.
Ecosystem services shown in blue. Physical processes related to salinity,

light and CO_2 availability shown in white within light-green boxes. Gene losses and gains associated with morphological and physiological processes shown in white within the dark-green box on the right.

Extended Data Table 1 | Genes involved in stomata development in *Z. marina* compared to other angiosperms

Gene Name	Symbol	<i>A. thaliana</i>	<i>O. sativa</i>	<i>S. polystachya</i>	<i>Z. marina</i>
Differentiation Genes					
SPEECHLESS	SPCH	At5g53210	Os02g15760 Os06g33450	Sp6G0039300	NF-1
MUTE	MUTE	At3g06120	Os05g51820		NF-1
FAMA	FAMA	At3g24140	Os05g50900		NF-1
SCREAM / ICE1	SCRM	At3g26744	Os11g32100	Sp4G0062100	Zm11g00170
SCREAM2 / ICE2	SCRM2	At1g12860	Os01g70310	Sp0G0129300	NF-1
FOUR LIPS	FLP	At1g14350			
MYB88	MYB88	At2g02820	Os07g43420	Sp0G0157900	NF-1
Spacing & Patterning Genes					
ERECTA	ER	At2g26330	Os06g10230	Sp15G0047400	Zm87g00130 Zm292g00090
ERECTA-LIKE1	ERL1	At5g62230			
ERECTA-LIKE2	ERL2	At5g07180	Os06g03970	Sp11G0029800	Zm85g01030
TOO MANY MOUTHS	TMM	At1g80080	Os01g43440	Sp18G0010300	NF-2
STOMATAL DENSITY & DISTRIBUTION	SDD1	At1g04110	Os03g04950	Sp1G0013100	NF-1
CO2 RESPONSE SECRETED PROTEASE	CRSP	At1g20160	Os09g30458	Sp3G0019800	Zm58g00010
EPIDERMAL PATTERNING FACTOR1	EPF1	At2g20875	Os04g54490 Os04g38470	Sp14G0058800 Sp15G0006400	NF-1 NF-1
EPIDERMAL PATTERNING FACTOR2	EPF2	At1g34245			
STOMAGEN/EPF-LIKE9	EPFL9	At4g12970	Os01g68598	Sp7G0057500	NF-1
CHALLAH/EPF-LIKE6	CHAL/EPFL6	At2g30370	Os01g60900 Os05g39880	Sp29G0014100	Zm270g00140
CHAL-LIKE1/EPF-LIKE5	CHALL1/EPFL5	At3g22820	Os03g06610	Sp2G0017500	Zm95g00050
CHAL-LIKE2/EPF-LIKE4	CHALL2/EPFL4	At4g14723	Os11g37190	Sp24G0023900	Zm289g00040
Polarity & Division Asymmetry Genes					
BREAKING OF ASYMMETRY IN THE STOMATAL LINEAGE	BASL	At5g60880	NC*	NC*	NC*
PANGLOSS1	PAN1†	At2g42290 At3g57830	Os08g39590	Sp12G0035200	Zm293g00080
PANGLOSS2	PAN2†	At4g20940	Os07g05190	Sp32G0009300 Sp0G0142000	Zm30g00950 Zm117g00680
POLAR LOCALIZATION DURING ASYMMETRIC DIVISION AND REDISTRIBUTION	POLAR	At4g31805	Os02g55190	Sp10G0014700	Zm16g01600
Cytokinesis Genes					
STOMATAL CYTOKINESIS DEFECTIVE 1	SCD1	At1g49040	Os01g39380	Sp21G0025200	Zm40g00290 Zm40g00310

The genes documented to be involved in stomatal development in *Arabidopsis*⁸⁶ were used as queries to find orthologues in rice and *Siproeta polystachya* (duckweed). See Supplementary Note 5.1, Supplementary Fig. 5.1 for sequence alignment and phylogenetic tree. NF-1, not found, supported by phylogeny; NF-2, not found, unambiguous reciprocal BlastP; NC, not conserved.

*BASL is not evolutionarily conserved, precluding the finding of its homologue in monocots, if it would exist.

†PAN genes have been searched for using the documented PAN1 and PAN2 genes from maize as baits.

Extended Data Table 2 | Ethylene-responsive transcription factor genes (ERF) in *Zostera marina*

Gene Family	<i>A. thaliana</i>	<i>O. sativa</i>	<i>S. polystachya</i>	<i>Z. marina</i>	Tissue expression in <i>Z. marina</i> (FPKM)				
ACS / ACSL					FFE	FFL	MF	R	L
1-aminocyclopropane-1-carboxylate synthase	AtACS1								
	AtACS2	OsACS2	NF	NF-1					
	AtACS6								
		OsACS5							
	AtACS7	OsACS6	NF	NF-1					
		OsACS7							
	AtACS4								
	AtACS5								
	AtACS8	OsACS1	Sp24g0002100	NF-1					
	AtACS9								
	AtACS11								
ACS-like	AtACS10								
	AtACS12	OsACS12	Sp1g0093100	Zm85g01020	8.7	16.7	16.3	11.3	18.1
AC0					FFE	FFL	MF	R	L
1-aminocyclopropane-1-carboxylate oxidase	AtACO1	Os06g37590 Os01g39860	Sp23g0011700	NF-1					
	AtACO2	Os02g53180							
	AtACO3	Os09g27750	NF-1	NF-1					
	AtACO4	Os09g27820							
	AtACO5	Os05g05680 Os11g08380	NF-1	NF-1					
ETR, ERS, EIN4					FFE	FFL	MF	R	L
Ethylene Receptors	AtETR1	Os03g49500	Sp6G0049300						
	AtERS1	Os05g06320	Sp22g0015200	NF-1					
	AtEIN4	Os04g08740							
	AtETR2	Os02g57530	Sp1g0021500 Sp23g0013000	NF-1					
	AtERS2	Os07g15540							
CTR1, EIN2 & Co					FFE	FFL	MF	R	L
Signaling genes and interacting partners	AtCTR1	Os02g32610 Os09g39320	Sp0g0009700	NF-1					
	AtEDR1	Os03g06410	NF-1	Zm289g00100	17.7	13.1	11.9	22.4	22.5
	AtEIN2	Os07g06130* Os03g49400*	Sp8g0029200	NF-1					
	AtRTE1	Os01g51430 Os05g46240	Sp14g0010800	NF-1					
	AtRTH	Os03g58520	Sp2g0051000	Zm159g00460	16.4	23.5	29.9	31.2	33
	AtRAN1	Os02g07630 Os06g45500	Sp8g0019500	Zm56g01580	10.8	16.7	15.2	72	35
	AtHMA5	Os04g46940 Os02g10290	Sp12g0033200	Zm25g00180	3.1	6	26.4	19.3	2.8
	AtEIN3	Os07g48630 Os03g20780	Sp3g0015900 Sp0g0106000	Zm44g00270 Zm140g00280	11 2	19.8 0.4	115 0.8	124 0	70 2.7
	AtEIL1	Os03g20790							
	AtEBF1	Os02g10700	Sp21g0000800						
	AtEBF2	Os06g40360	Sp27g0021100 Sp27g0021200	NF-1					
	AtXRN4	Os03g58060	Sp2g0012600	Zm177g00170	15.6	10.3	13.4	16.9	20.2

MF, male flowers; FFE, female flowers early; FFL, female flowers late; R, roots; L, leaves; NF-1, not found as supported by reciprocal Blast and phylogeny. See Supplementary Note 6.1, Supplementary Fig. 6.1 for sequence alignment and phylogenetic tree. Grey indicates genes not involved in ethylene biosynthesis and signal pathways but strongly co-expressed, indicative of multiple functions.

Extended Data Table 3 | Genes involved in pollen development of *Z. marina* compared to other angiosperms

Gene Name	Symbol	<i>A. thaliana</i>	<i>O. sativa</i>	<i>S. polystachya</i>	<i>Z. marina</i>	Tissue expression in <i>Z. marina</i> (FPKM)				
						FFE	FFL	MF	R	L
ACYL-COA SYNTHETASE 5	ACOS5	At1g62940	Os04g24530	Sp12g0064500	NF-1					
POLYKETIDE SYNTHASE A (LAP6)	PKSA (LAP6)	At1g02050	Os10g34360	Sp16g0013800	NF-1					
POLYKETIDE SYNTHASE B (LAP5)	PKSB (LAP5)	At4g34850	Os07g22850	Sp1g0062300	NF-1					
LESS ADHERENT POLLEN 3	LAP3	At3g59530	Os03g15710	Sp16g0030000	NF-1					
TETRAKETIDE α -PYRONE REDUCTASE 1	TKPR1 (DRL1)	At4g35420	Os08g40440	Sp10g0016700	NF-1					
TETRAKETIDE α -PYRONE REDUCTASE 2	TKPR2 (CCRL6)	At1g68540	Os01g03670	NF	NF-1					
CYTOCHROME P450 704B1/2	CYP704B1(CYP704B23)	At1g69500	Os03g07250	Sp2g0036600	Zm149g00275	NA	NA	NA	NA	NA
TYPE III LIPID TRANSFER PROTEINS	LTP3	At5g62080 At5g07230 At5g52160	Os08g43290 Os09g35700	Sp16g0007800	NF-1					
GA-regulated Myb-like Transcription Factor	GAMYB (MYB65 MYB33)	At3g11440 At5g06100	Os01g59660	Sp22g0020200	Zm6g00090	8.8	6.6	26.9	3.1	4.3
FACELESS POLLEN-1, ECERIFERUM 3	FLP1 (ERC3, WAX2)	At5g57800	Os09g25850 Os02g08230 Os06g44300	Sp5g0009000	NF-1					
INAPERTURATE POLLEN 1	INP1	At4g22600	Os02g44250	Sp13g0036900	NF-2					
GLYCOSYLTRANSFERASE 1	GT1	At1g19710 At1g75420	Os01g15780	Sp14G0031700	Zm69g00440	10.1	8.1	41.7	224.1	62.4
CYSTEINE ENDOPEPTIDASE 1	CEP1	At5g50260	Os08g44270 Os11g14900	Sp4g0036900 Sp11g0013300	NF-1					
MALE STERILE 188	MS188 (MYB80)	At5g56110	Os04g39470	Sp4g0087200	Zm262g00100	5.9	4.6	42.5	9.4	2.8
Fatty Acyl Thioesterase B Glycosyl transferase family GT31	FATB	At1g08510	Os06g05130	Sp21g0008900	Zm1g01370	115.1	104.8	213.7	129.3	124.3
NO EXINE FORMATION	B3GALT7	At1g77810	Os02g35870	Sp4g0008000	Zm155g00180	4.4	1.4	354.3	10.5	2.2
	NEF1	At5g13390	Os11g32470	Sp9g0054300	Zm5g01490	28.4	23.6	14.7	25.3	18.8

The five genes encoding proteins associated on the ER-located sporopollenin metabolon in *Arabidopsis*⁸⁷ are highlighted in grey. The genes documented to be involved in pollen development in *Arabidopsis* or in rice were used as queries to find orthologues. MF, male flowers; FFE, female flowers early; FFL, female flowers late; R, roots; L, leaves; NF-1, not found, supported by reciprocal Blast and phylogeny; NF-2, not found, single copy gene; amb, ambiguous with homologues too similar to point to a specific orthologue. See Supplementary Note 11.1, Supplementary Fig. 11.1 for sequence alignment and phylogenetic tree; Supplementary Table 11.1 for complete gene list.

Polygenic evolution of a sugar specialization trade-off in yeast

Jeremy I. Roop¹, Kyu Chul Chang^{2,3} & Rachel B. Brem^{1,3}

The evolution of novel traits can involve many mutations scattered throughout the genome^{1,2}. Detecting and validating such a suite of alleles, particularly if they arose long ago, remains a key challenge in evolutionary genetics^{1–3}. Here we dissect an evolutionary trade-off of unprecedented genetic complexity between long-diverged species. When cultured in 1% glucose medium supplemented with galactose, *Saccharomyces cerevisiae*, but not *S. bayanus* or other *Saccharomyces* species, delayed commitment to galactose metabolism until glucose was exhausted. Promoters of seven galactose (*GAL*) metabolic genes from *S. cerevisiae*, when introduced together into *S. bayanus*, largely recapitulated the delay phenotype in 1% glucose–galactose medium, and most had partial effects when tested in isolation. Variation in *GAL* coding regions also contributed to the delay when tested individually in 1% glucose–galactose medium. When combined, *S. cerevisiae* *GAL* coding regions gave rise to profound growth defects in the *S. bayanus* background. In medium containing 2.5% glucose supplemented with galactose, wild-type *S. cerevisiae* repressed *GAL* gene expression and had a robust growth advantage relative to *S. bayanus*; transgenesis of *S. cerevisiae* *GAL* promoter alleles or *GAL* coding regions was sufficient for partial reconstruction of these phenotypes. *S. cerevisiae* *GAL* genes thus encode a regulatory program of slow induction and avid repression, and a fitness detriment during the glucose–galactose transition but a benefit when glucose is in excess. Together, these results make clear that genetic mapping of complex phenotypes is within reach, even in deeply diverged species.

A central goal of evolutionary genetics is to understand how organisms acquire phenotypic novelties. Such traits, if they have evolved over long timescales, can have a genetic basis quite distinct from those arisen more recently⁴. In landmark cases, single genes underlying species differences have been pinpointed and validated⁵, but the polygenic architecture of ancient traits has remained a mystery.

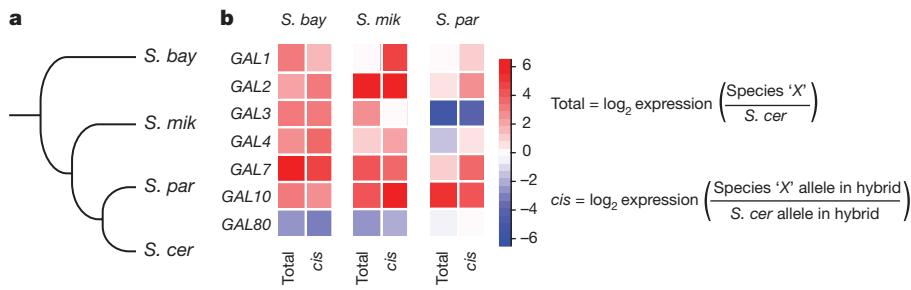


Figure 1 | Polygenic *cis*-regulatory evolution among yeast species in galactose metabolic genes. **a**, Phylogenetic tree of *Saccharomyces* species studied here¹⁴. *S. bay*, *S. bayanus*; *S. mik*, *S. mikatae*; *S. par*, *S. paradoxus*; *S. cer*, *S. cerevisiae*. **b**, Each cell reports expression, as a ratio between the indicated species and *S. cerevisiae*, of the indicated galactose metabolism

In hybrids formed by mating *S. cerevisiae* with other *Saccharomyces* species⁶, we noted a pattern of coherent *cis*-regulatory variation in the seven genes of the galactose metabolic pathway. During growth in medium with glucose as the sole carbon source, the *S. cerevisiae* allele at each *GAL* gene conferred low expression relative to other *Saccharomyces*, except for the repressor *GAL80*, at which the *S. cerevisiae* allele drove expression up (Fig. 1b). Likewise, purebred *S. cerevisiae* expressed *GAL* effectors at low levels in glucose, and *GAL80* at high levels, relative to other species (Fig. 1b and ref. 7). *S. paradoxus*, the sister species to *S. cerevisiae*, had an intermediate expression phenotype (Fig. 1b). Thus, the *S. cerevisiae* *GAL* program is one of heightened glucose repression relative to other species, as a product of *cis*-regulatory changes at the five loci that encode the seven *GAL* genes. Because such a pattern is unlikely under neutrality⁸, these data raised the possibility that selective pressure on the *GAL* pathway had changed along the *S. cerevisiae* lineage.

In *S. cerevisiae*, pre-expression of metabolic genes in glucose medium can boost fitness upon a switch to other carbon sources^{9–11}. We therefore expected that *GAL* expression divergence in glucose could have phenotypic correlates in other conditions. Culturing cells in 1% glucose–galactose medium, we observed a qualitative distinction between species (Fig. 2a). In *S. cerevisiae*, growth was retarded by a diauxic lag midway through the time course, reflecting the expected delay in assembling galactose metabolic machinery once glucose is exhausted^{9,10,12}. In more distantly related yeasts, we observed no lag in 1% glucose–galactose medium supplemented with galactose (Fig. 2a, b), although *S. paradoxus* had a modest lag (Fig. 2a, b) that echoed its intermediate regulatory phenotype (Fig. 1b). Glucose mixtures with maltose and raffinose engendered a lag in all members of the clade (Extended Data Fig. 1). *S. cerevisiae* strains from distinct populations all exhibited a lag in glucose–galactose cultures (Fig. 2c). These data highlight *S. cerevisiae* as an extreme among *Saccharomyces* with respect to two attributes of galactose metabolism: reduced *GAL*

¹Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, California 94720, USA. ²Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, California 94720, USA. ³Buck Institute for Research on Aging, Novato, California 94945, USA.

gene during culture in glucose medium⁶. Total expression measured in purebred species; *cis*, expression from the indicated species' allele in a diploid hybrid between this species and *S. cerevisiae*, reflecting effects of *cis*-regulatory divergence.

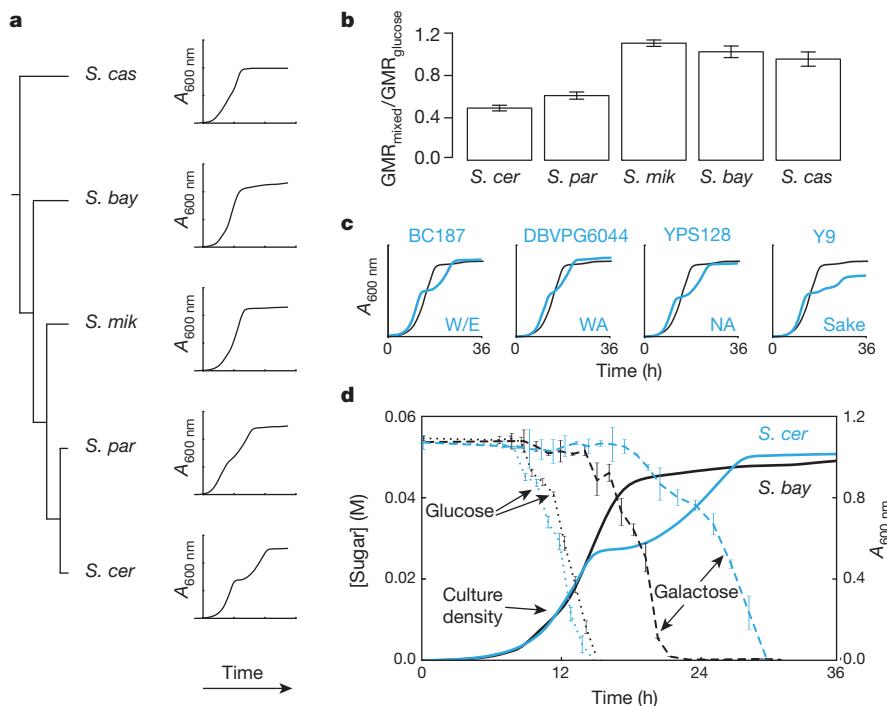


Figure 2 | Diauxic lag, in 1% glucose–galactose medium, is conserved within *S. cerevisiae* and divergent among species. **a**, Growth of *Saccharomyces* type strains inoculated into medium containing 1% glucose and 1% galactose ($n=6$). *S. cas*, *S. castellii*. **b**, Each bar reports the geometric mean of the growth rate (GMR) of the indicated species from the time course in **a**, normalized to the analogous quantity in glucose medium. **c**, Growth of *S. cerevisiae* isolates (blue) from the indicated populations (W/E, Wine/European; WA, West African; NA, North

gene expression during growth in pure glucose, and diauxic lag in 1% glucose–galactose.

To dissect further the divergence in galactose metabolic behaviours, we focused on a comparison of *S. cerevisiae* with its distant relative *S. bayanus* var. *uvarum* (*S. bayanus*). In 1% glucose–galactose medium, both species initially metabolized glucose with similar rates, indicating that neither used the sugars simultaneously (Fig. 2d). In *S. bayanus* cultures, galactose consumption began at a point just before the complete exhaustion of glucose. For *S. cerevisiae*, glucose exhaustion triggered the diauxic lag, during which galactose levels in its culture medium were largely unchanged. After the lag, with the eventual resumption of log-phase growth by *S. cerevisiae*, galactose levels finally dropped (Fig. 2d). These results implicate the transition between glucose and galactose metabolism as a nexus of phenotypic differences between the species.

For direct tests of the phenotypic impact of divergence at the *GAL* genes, we replaced *GAL* gene sequences in one species by those of the other at the endogenous loci (Fig. 3a). In a first investigation of *GAL* promoters, *S. cerevisiae* alleles of the regions upstream of *GAL1*, *GAL3*, *GAL4*, and *GAL10* were each sufficient for a partial gain in diauxic lag in *S. bayanus*, in 1% glucose–galactose medium (Fig. 3b, c). Control experiments established the inverse effect of *S. bayanus* *GAL* promoter alleles, reducing lag in the *S. cerevisiae* background (Extended Data Fig. 2).

We next aimed at a more complete reconstruction of *S. cerevisiae*-like galactose metabolic behaviours, which we inferred to be derived, in *S. bayanus* as a representative of the likely ancestral state. An *S. bayanus* strain harbouring all seven *GAL* promoters from *S. cerevisiae* recapitulated 69% of the lag phenotype of the *S. cerevisiae* parent (Fig. 3b, c), with *GAL* gene expression peaking at the same time point as that of wild-type *S. cerevisiae* and at similar amplitude (Fig. 3d). Comparison with the sum of lag effects from individual promoter swaps revealed

American) and the *S. bayanus* type strain (black), inoculated into medium containing 1% glucose and 1% galactose ($n=6$). **d**, Growth (solid lines, $n=4$) of *S. cerevisiae* and *S. bayanus* inoculated into medium containing 1% glucose and 1% galactose, and medium concentrations of glucose and galactose (dotted and broken lines, respectively; $n=2$ biological replicates, each comprising 3 technical replicates). Error bars, s.e.m. Each set of data is representative of the results of two independent experiments.

negative epistasis in the seven-promoter replacement strain (Fig. 3c), and in strains harbouring intermediate *S. cerevisiae* promoter combinations (Extended Data Fig. 3).

Transgenesis of individual *S. cerevisiae* *GAL* coding regions was also sufficient for a partial lag in *S. bayanus*, in the case of *GAL1*, *GAL2*, *GAL3*, *GAL4*, *GAL7*, and *GAL10* (Fig. 3b, c). Swaps of *GAL* promoter-coding fusions revealed negative epistasis at *GAL1* and *GAL3*: for these genes, the sum of phenotypes from the respective promoter and coding transgenics was far more dramatic than the effect of the promoter-coding fusion (Fig. 3c). Combining all seven *S. cerevisiae* *GAL* coding or promoter-coding regions in *S. bayanus*, we observed an exaggerated, long-term growth delay in 1% glucose–galactose medium, distinct from the temporary lag of wild-type strains and promoter transgenics (Fig. 3b). This defect reflected dysfunction of multiple modules of the *S. cerevisiae* *GAL* pathway in *S. bayanus*, as it could be elicited by just the two regulators Gal3 and Gal4 swapped from *S. cerevisiae*, or just *S. cerevisiae* alleles of the enzymes Gal1, Gal7, and Gal10 (Extended Data Fig. 3). Coding and promoter-coding swap strains did ultimately resume active growth (Fig. 3b) and metabolize galactose from mixed-sugar medium (Extended Data Fig. 4), and their *GAL* expression induction was markedly delayed (Fig. 3d). These strains also grew poorly in pure galactose medium (Extended Data Fig. 5). Together, our data make clear that diauxic lag in 1% glucose–galactose medium can be largely recapitulated by divergent *GAL* gene promoters; *GAL* protein alleles from *S. cerevisiae* make a partial contribution to lag when tested in isolation and, when combined in *S. bayanus*, confer growth defects far exceeding those of either wild-type.

In light of the conservation of diauxic lag across *S. cerevisiae* (Fig. 2c), we hypothesized that this species had maintained its divergent galactose metabolic behaviour on the basis of a fitness benefit. Among the potential mechanisms for such an advantage, we focused on the possibility that as *S. cerevisiae* represses *GAL* genes in glucose-replete

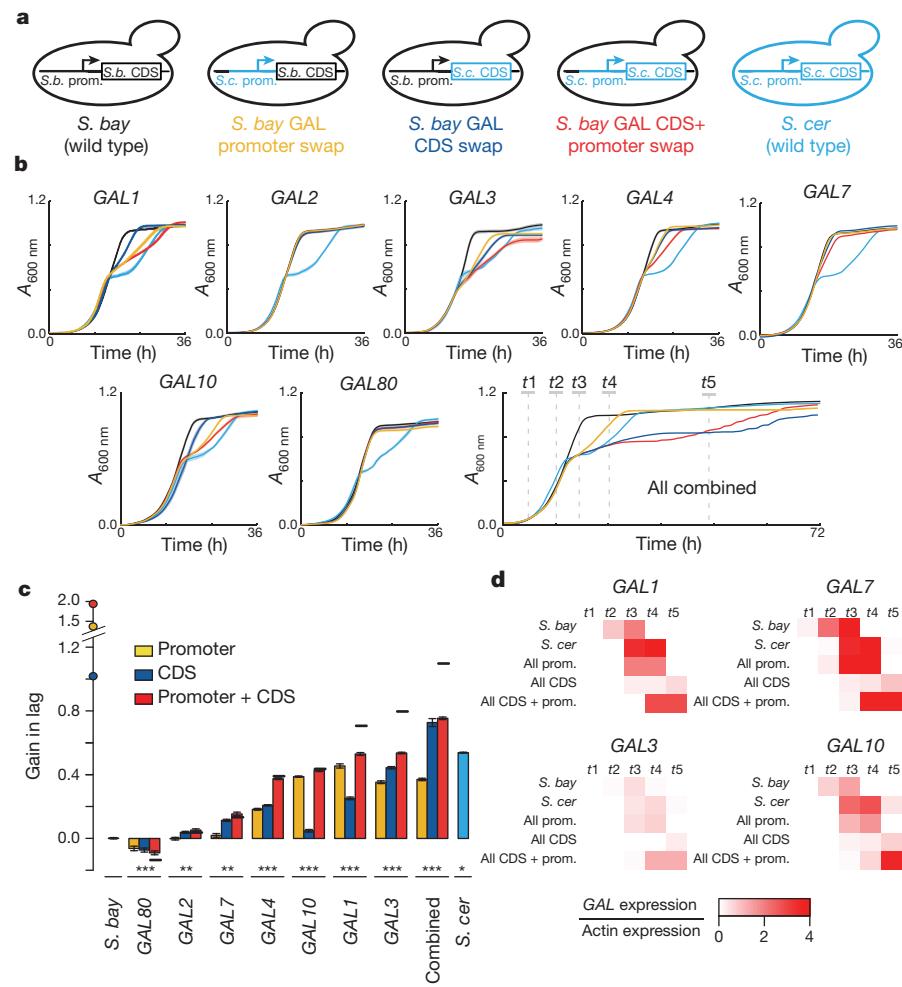


Figure 3 | *S. cerevisiae* alleles of *GAL* genes confer diauxic lag in 1% glucose-galactose medium. **a**, Replacement of *S. cerevisiae* *GAL* sequences into *S. bayanus* at the endogenous loci. **b**, Growth of *S. bayanus* harbouring *S. cerevisiae* alleles of a single *GAL* gene, or of all seven genes, inoculated into medium containing 1% glucose and 1% galactose ($n \geq 12$). **c**, Each bar reports the ratio of the GMR over the time course of the indicated strain in **b** to that of wild-type *S. bayanus*, subtracted from 1; negative values are GMRs faster than wild-type. Error bars, s.e.m.

conditions (Fig. 1), it avoids the liability of expressing unused proteins and enables rapid growth^{10,13}. When cultured in 2.5% glucose medium also containing galactose, wild-type *S. cerevisiae* exhibited a 10% faster growth rate (Fig. 4a, b), and fourfold to ninefold lower expression of *GAL* enzymes (Fig. 4c), than *S. bayanus*. Both species metabolized glucose almost exclusively across the time course (Fig. 4d, e). Replacement of all seven *S. bayanus* *GAL* promoters with *S. cerevisiae* alleles recapitulated the program of low *GAL* gene expression (Fig. 4c), and conferred a growth rate halfway between those of the wild-type species (Fig. 4a, b), in 2.5% glucose medium supplemented with galactose. The *S. bayanus* strain harbouring all seven *S. cerevisiae* *GAL* coding regions also expressed *GAL* genes at low levels (Fig. 4c), which mirrored this strain's exaggerated delay in *GAL* gene induction (Fig. 3b-d), and was associated with a partial growth benefit (Fig. 4a, b). Promoter-coding replacement conferred no additional phenotype over and above the effects of transgenesis of either region type alone (Fig. 4a-c). We conclude that *S. cerevisiae* *GAL* promoters, by shutting down expression of the galactose metabolic pathway, are adaptive in conditions of abundant glucose, and this program can be phenocopied by *S. cerevisiae* *GAL* protein alleles in *S. bayanus*. Sequence analyses revealed a high ratio of inter-specific divergence to intra-species polymorphism in *GAL* gene promoters, and not in *GAL*

Asterisks, significant differences ($P < 0.001$, Wilcoxon rank-sum) from wild-type *S. bayanus*. Also shown are expected phenotypes of promoter-CDS transgenics for a single gene (horizontal lines) or seven-locus transgenics (circles on y axis), under an additive model of contributions from the regions combined in the respective strains. **d**, *GAL* gene expression at time points indicated in the final panel of **b** ($n \geq 2$ biological replicates, each comprising 3 technical replicates). Each set of data is representative of the results of two independent experiments.

coding regions (Extended Data Tables 1–3), suggestive of a history of directional evolution at these loci.

In this work, we have dissected glucose-specialist phenotypes that distinguish *S. cerevisiae* from other members of the *Saccharomyces* clade. *S. cerevisiae* is reluctant to transition from glucose to galactose metabolism, and has a growth advantage in a high-glucose environment. Additionally, the *S. cerevisiae* program confers an increase in biomass accumulation during growth in pure galactose (Extended Data Fig. 5c) and could be beneficial when glucose availability fluctuates rapidly¹¹. As *S. cerevisiae* alleles of *GAL* gene promoters are largely sufficient for this family of traits, they may have served as an easily evolvable, and probably adaptive, origin of these characters. By contrast, the *S. cerevisiae* *GAL* proteome, which confers synthetic growth defects in modern-day *S. bayanus*, may have evolved slowly over a rugged fitness landscape, under distinct forces or at a different period. Such a model would dovetail with the *cis*-regulatory basis of a related, but genetically simple, galactose metabolism trait that evolved more recently between yeasts¹³. For any suite of divergent regulatory regions, observing *cis*-acting effects on gene expression can open a first window onto their phenotypic relevance and that of the gene products they control. With this strategy, evolutionary biologists need not be limited by polygenicity in the mapping of genotype to phenotype, even between long-diverged species.

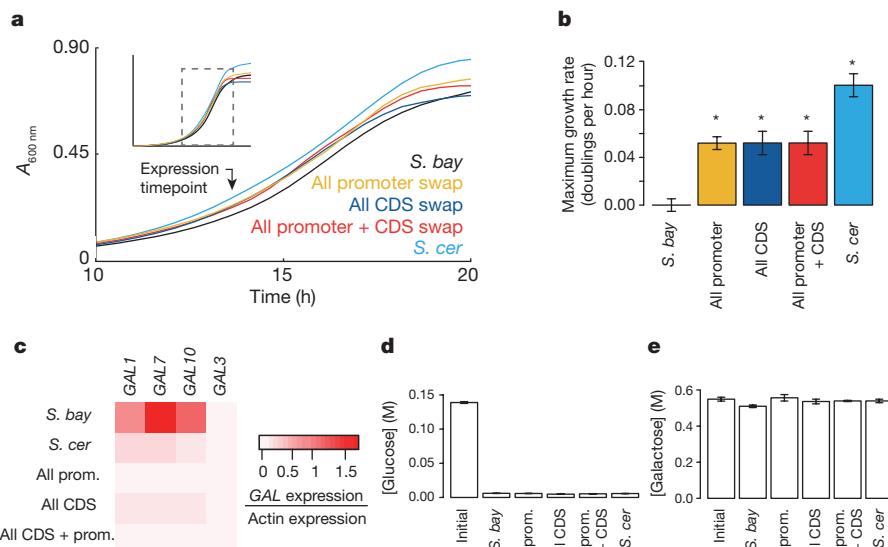


Figure 4 | *S. cerevisiae* *GAL* alleles confer a fitness advantage in 2.5% glucose medium supplemented with galactose. a, Growth of *S. bayanus* strains harbouring *S. cerevisiae* alleles of all seven *GAL* genes, and the wild-type species, inoculated into medium containing 2.5% glucose and 10% galactose. Inset shows the complete time course from which the main figure shows a narrower time window ($n \geq 136$, multiple independent experiments pooled). **b**, Each bar reports the difference in maximum growth rate between the indicated species and wild-type *S. bayanus* over the

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 March; accepted 18 December 2015.

Published online 10 February 2016.

1. Orr, H. A. The genetic theory of adaptation: a brief history. *Nature Rev. Genet.* **6**, 119–127 (2005).
2. Rockman, M. V. The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution* **66**, 1–17 (2012).
3. Pritchard, J. K., Pickrell, J. K. & Coop, G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr. Biol.* **20**, R208–R215 (2010).
4. Savolainen, O., Lascoux, M. & Merilä, J.; Savolainen. Ecological genomics of local adaptation. *Nature Rev. Genet.* **14**, 807–820 (2013).
5. Nadeau, N. J. & Jiggins, C. D. A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations. *Trends Genet.* **26**, 484–492 (2010).
6. Schraiber, J. G., Mostovoy, Y., Hsu, T. Y. & Brem, R. B. Inferring evolutionary histories of pathway regulation from transcriptional profiling data. *PLoS Comput. Biol.* **9**, e1003255 (2013).
7. Caudy, A. A. et al. A new system for comparative functional genomics of *Saccharomyces* yeasts. *Genetics* **195**, 275–287 (2013).
8. Bullard, J. H., Mostovoy, Y., Dudoit, S. & Brem, R. B. Polygenic and directional regulatory evolution across pathways in *Saccharomyces*. *Proc. Natl. Acad. Sci. USA* **107**, 5058–5063 (2010).
9. Venturelli, O. S., Zuleta, I., Murray, R. M. & El-Samad, H. Population diversification in a yeast metabolic program promotes anticipation of environmental shifts. *PLoS Biol.* **13**, e1002042 (2015).

time course of the indicated strain in **a**. Error bars, s.e.m.; asterisks indicate rates significantly different ($P < 1 \times 10^{-8}$, Wilcoxon rank-sum) from wild-type *S. bayanus*. **c**, Each row reports *GAL* gene expression of the indicated strain at time points indicated by the arrow in **a** ($n \geq 2$ biological replicates, each comprising 3 technical replicates). **d**, **e**, In each panel, the first bar reports sugar concentration in medium before inoculation; the remaining bars report sugar concentrations in medium after the indicated time course in **a** ($n = 3$ biological replicates, each comprising 3 technical replicates).

10. Wang, J. et al. Natural variation in preparation for nutrient depletion reveals a cost-benefit tradeoff. *PLoS Biol.* **13**, e1002041 (2015).
11. New, A. M. et al. Different levels of catabolite repression optimize growth in stable and variable environments. *PLoS Biol.* **12**, e1001764 (2014).
12. De Deken, R. H. The Crabtree effect: a regulatory system in yeast. *J. Gen. Microbiol.* **44**, 149–156 (1966).
13. Peng, W., Liu, P., Xue, Y. & Acar, M. Evolution of gene network activity by tuning the strength of negative-feedback regulation. *Nature Commun.* **6**, 6226 (2015).
14. Scannell, D. R. et al. The awesome power of yeast evolutionary genetics: new genome sequences and strain resources for the *Saccharomyces* *sensu stricto* genus. *G3* **1**, 11–25 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by National Institutes of Health GM087432 to R.B.B. and a Hellman Graduate Fellowship from the University of California Berkeley to J.I.R. We thank A. Arkin for his advice and resources; J. Rine for yeast strains; O. Venturelli for technical expertise; and A. Flamholz, J. Schraiber, and P. Shih for discussions.

Author Contributions J.I.R. and R.B.B. designed experiments, J.I.R. and K.C.C. conducted experiments, J.I.R. analysed data, and J.I.R. and R.B.B. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.B.B. (rbrem@buckinstitute.org).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Yeast strains. Strains used in this study are listed in Supplementary Table 1. Abbreviations in figures and tables are as follows: *S. cer*, *Saccharomyces cerevisiae*; *S. par*, *Saccharomyces paradoxus*; *S. mik*, *Saccharomyces mikatae*; *S. bay*, *Saccharomyces bayanus*; *S. cas*, *Saccharomyces castellii*. Allele-swap strains constructed in haploid *S. bayanus* JRY294 or JRY296 (isogenic MAT α and MAT α derivatives of type strain CBS7001) used the MIRAGE method¹⁵ with several modifications as follows. A 1.7-kilobase region containing the *K. lactis* URA3 coding sequence and regulatory region was amplified from the pCORE-UH plasmid¹⁶ and used for each half of the inverted repeat. The *S. cerevisiae* *GAL* region to be swapped in was attached to one half of the inverted repeat cassette by overlap extension PCR, after which the two halves of the final cassette were ligated together. Owing to the different sizes of the two halves of the inverted repeat cassette, a second restriction digestion step as described in ref. 15 was not necessary to remove non-desired ligation products. Transformation with the cassette, followed by confirmation of positive transformants and plating onto 5-FOA medium, resulted in excision of the inverted repeat from the target genome, leaving behind a marker-less allele swap at the locus. Sanger sequencing was used to verify the correct nucleotide sequence of each swapped allele. The *S. cerevisiae* allele of each promoter and CDS was amplified from genomic DNA of YHL068 (ref. 6). Promoter, coding, and promoter-coding fusion swap strains were engineered by replacing 600 base pairs (bp) of intergenic region directly upstream of the CDS, the CDS, or these two regions combined, respectively, in *S. bayanus* with orthologous regions from *S. cerevisiae*. For allele swaps in the *S. cerevisiae* background in Extended Data Fig. 2, 720 bp of the region between the *GAL1* and *GAL10* open reading frames were amplified from *S. bayanus* strain CBS7001 and used to construct a MIRAGE cassette as above, and transformed into *S. cerevisiae* strain JRY313 (isogenic MAT α derivative of BY4743) and selected as above. For each transgenic, two or more independent transformants were used as replicates for growth profiling and sugar concentration measurements. Combining unlinked allele swaps into a single genome was accomplished by single-cell mating of single-locus swaps, followed by sporulation, tetrad dissection, and diagnostic PCR to identify segregant colonies with the allele combinations of interest. *S. cerevisiae* alleles of *GAL1*, *GAL7* and *GAL10* were combined in the *S. bayanus* background by successive allele swap transformations.

Growth curves and quantification. All growth experiments were conducted at 26 °C in YP media (2% bacto-peptone, 1% yeast extract) supplemented with various carbon sources as follows. Experiments measuring diauxic lag in 1% glucose medium supplemented with galactose used medium containing 1% glucose and 1% galactose. Experiments measuring growth profiles in other non-galactose carbon sources used media containing 1% glucose and 1% of the secondary carbon source as indicated. Experiments measuring maximum growth rates in high-glucose media containing galactose used medium containing 2.5% glucose and 10% galactose. Experiments measuring growth profiles in pure galactose medium used media containing 2% galactose.

Growth time courses were performed as follows. Strains were grown in YP containing 2% galactose (Fig. 4 and Extended Data Fig. 5) or 2% glucose (all other figures) for 24 h with shaking at 200 r.p.m. Each strain was then back-diluted into the same medium to an absorbance (*A*) of 0.1 and grown for an additional 6 h. These log-phase cultures were then back-diluted to an absorbance of 0.02 in a 96-well plate containing 150 μ l of YP with the appropriate amount of a given carbon source. Plates were covered with a gas-permeable membrane, placed in a Tecan F200 plate reader and incubated with orbital shaking for the duration of each experiment. Measurements of $A_{600\text{ nm}}$ were made every 30 min.

Gain in lag was calculated from growth curves as $(1 - (\text{GMR of a given strain}) / (\text{GMR of } S. bayanus))$. GMR was calculated as in ref. 11, with the following differences. A window of 0.1–0.8 absorbance units was used for quantification in all figures apart from Fig. 2b. In Fig. 2b, GMR was calculated within a window bounded by 20% and 80% of the maximum final yield attained during the time course for each species. Maximum growth rate was calculated as in ref. 10 except that a window of 0.01–0.3 absorbance units was used and a geometric mean of growth rates was calculated. Final growth yield was calculated as the difference between initial and final $A_{600\text{ nm}}$ measurements as in ref. 17.

Replication schemes and analysis were as follows. To enable qualitative comparisons among species in Fig. 2a–c, on a given 96-well plate, six biological replicate cultures of each strain were assayed. On each growth plate, replicate fitness values greater than two standard deviations from the mean of fitness values for that strain were considered artefacts of technical error and discarded. Displayed data for a given strain are the results of growth measurements from one plate and are representative of at least two plates cultured on different days. Experiments

in Fig. 2d were as in Fig. 2a–c except that four biological replicate cultures were measured.

To enable highly powered quantitative comparisons among strains of growth in 1% glucose–galactose medium in Fig. 3b, c, 12 biological replicate cultures of each transgenic strain were assayed across several plates and days, in each case alongside replicates of wild-type *S. bayanus* and *S. cerevisiae*. The growth rate of a given strain measured on a given plate was normalized to the value for wild-type *S. bayanus* on that plate, and these normalized measurements were then averaged across plates. Because we included wild-type strains on each plate, their growth measurements as displayed in Fig. 3b, c are averages over 78 and 54 replicates of *S. bayanus* and *S. cerevisiae*, respectively. Artefact filtering was as above. Differences in growth among strains were assessed for statistical significance by a two-sided Wilcoxon rank-sum test and a Bonferroni correction was applied in instances of multiple tests.

To enable highly powered quantitative comparisons among strains of growth in medium containing 2.5% glucose and 10% galactose in Fig. 4a, b, we assayed growth of 250, 170, 204, 136, and 192 biological replicate cultures of wild-type *S. bayanus*, wild-type *S. cerevisiae*, the combinatorial promoter transgenic strains, the combinatorial promoter-coding transgenic strains, and the combinatorial coding transgenic strains, respectively, across several plates and days. As above, growth measurements for each strain in turn assayed on a given plate were normalized to the wild-type *S. bayanus* cultured on that plate, and normalized growth rate measurements were combined across plates. Artefact filtering and statistical testing were as above.

Experiments in Extended Data Fig. 1 were as in Fig. 2a–c. Experiments for Extended Data Fig. 2 were as in Fig. 3 except that 12 replicate cultures were measured for each strain. Experiments for Extended Data Fig. 3 were as in Fig. 3b, c. Experiments for Extended Data Fig. 4 were as in Fig. 2a–c.

Sugar measurements. For growth time-course experiments in which sugar concentration was measured, an appropriate number of wells (see below) were inoculated into a 96-well plate and cultured in the Tecan F200 plate reader as above, such that media from at least two replicate wells could be harvested at each time point of interest and at least four replicate cultures of each strain would remain untouched for growth curve analysis. Samples were taken for sugar measurement by cutting the membrane covering the 96-well plate with a razor and extracting all 150 μ l of cell culture in a given well. Care was taken only to puncture the membrane above the harvested well such that adjacent wells were not affected. Cells and debris were pelleted from the sampled culture by brief centrifugation and the supernatant was extracted for quantification of glucose and galactose. Glucose was measured using the GlucCell glucose monitoring system (Chemglass Life Sciences). Galactose was measured with the Amplex Red Galactose Oxidase assay kit (Molecular Probes, Life Technologies). For galactose measurements, a Tecan Safire plate reader was used to quantify fluorescence and the relationship between fluorescence and galactose concentration was determined using a standard curve.

Data in Fig. 2d were obtained by sampling two biological replicate cultures at each time point. Data in Fig. 4d, e and Extended Data Fig. 4 were obtained by sampling three biological replicate cultures both at the start and at the endpoints of the growth time course. For all experiments, three technical replicates were assayed for each biological replicate culture sampled, and mean values are reported. All data are representative of two identical experiments conducted on different days.

Quantitative PCR. Time point samples for quantitative PCR analysis were obtained from cultures analogously to those obtained for sugar consumption quantification detailed above. Between 2 and 15 replicate wells were harvested at each time point and pooled to have sufficient biological material for RNA isolation. RNA was isolated using an RNeasy mini kit (Qiagen) and complementary DNA was synthesized using SuperScript III (Life Technologies). DyNAmo HS SYBR green (Thermo Scientific) was used for quantitative PCR and all quantification was done on a CFX96 machine (BioRad). Gene expression levels relative to *ACT1* were calculated using the $2^{-\Delta\Delta C_t}$ method¹⁸. Three technical replicates per biological sample were assayed, and mean values are reported. Primer sequences and cycling times are in Supplementary Table 2.

Sequence analyses. Custom Python scripts were used to extract coding sequences and 600 bp promoter regions for type strains of *S. paradoxus*, *S. mikatae*, and *S. bayanus*¹⁴ for each gene that had an orthologue in each of the five *Saccharomyces sensu stricto* species as reported in ref. 14. *S. cerevisiae* population sequences were downloaded from the following sources: YJM978, UWOPS83-787, Y55, UWOPS05-217.3, 273614N, YS9, BC187, YPS128, DBVPG6765, YJM975, L1374, DBVPG1106, K11, SK1, 378604X, YJM981, UWOPS87-2421, DBVPG1373, NCYC3601, YPS606, Y12, UWOPS05-227.2, and YS2 from <http://www.yeastrc.org/g2p/home.do>; Sigma1278b, ZTW1, T7, and YJM789 from <http://www.yeast-genome.org/>; and RM11 from http://www.broadinstitute.org/annotation/genome/Saccharomyces_cerevisiae. *S. cerevisiae* sequences were aligned to each of the other

three species in turn using FSA (ref. 19), with the ‘–nucprot’ option for the coding sequence alignments. For each set of species alignments, nucleotide replacement and polymorphic sites were tabulated using Polymorphorama²⁰ for coding regions and custom Python scripts for promoter regions.

Sequence analyses of the seven genes of the *GAL* pathway (*GAL1*, *GAL2*, *GAL3*, *GAL4*, *GAL7*, *GAL10*, and *GAL80*) using *S. paradoxus*, *S. mikatae*, or *S. bayanus* as an outgroup were done as follows. For a given species comparison, we first calculated the neutrality index NI_{TG} statistic²¹ for the promoter regions of the *GAL* genes using synonymous sites in the downstream gene as putative neutral sites, as

$$\frac{\sum D_{si}P_{ri} / (P_{si} + D_{si})}{\sum P_{si}D_{ri} / (P_{si} + D_{si})}$$

where i counts genes of the group, D_s and D_r denote the number of divergent synonymous and divergent promoter sites respectively, and P_s and P_r denote the number of polymorphic synonymous and polymorphic promoter sites respectively.

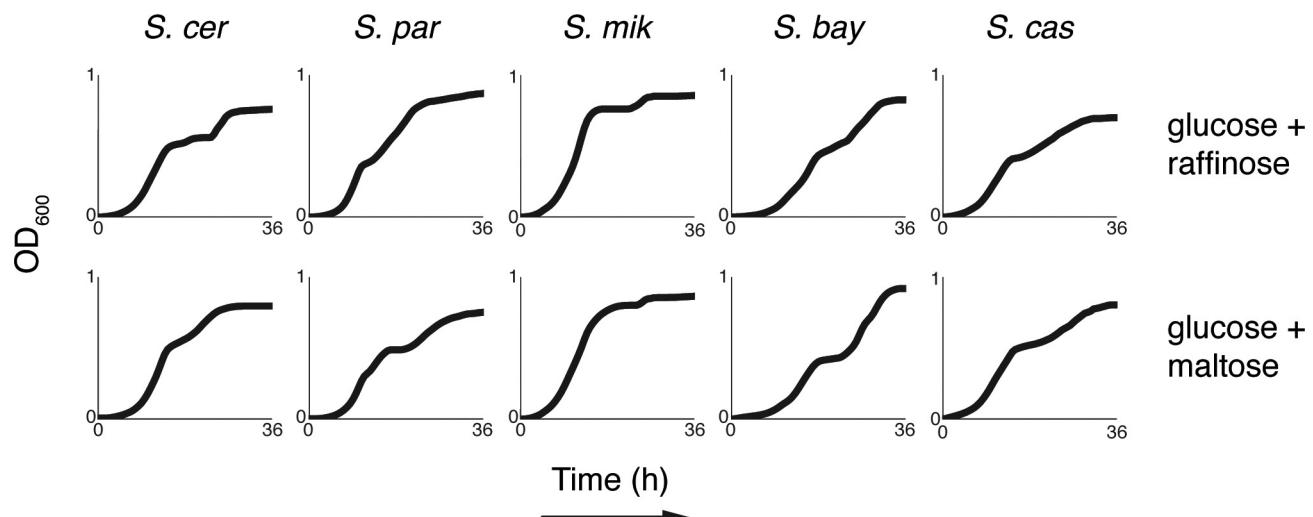
Minor alleles with a frequency of less than 0.15 were ignored^{22,23} and all site counts were corrected for multiple hits using a Jukes–Cantor model. In computing NI_{TG} for the seven genes of the *GAL* pathway, we considered that since *GAL1* and *GAL10* are adjacent genes on chromosome II sharing a 662 bp promoter region, simply counting sites within 600 bp promoter regions for each gene separately would have resulted in double-counting of 532 bp. To avoid this, we aligned the 662 bp intergenic region of these two genes and considered it one locus in the NI_{TG} calculation. Synonymous sites for this region were taken as the sum of such sites in both the *GAL1* and *GAL10* coding sequences. To evaluate significance, we generated 10,000 randomly chosen groups of six genes each and computed NI_{TG} across the promoters of each such null group as above. The empirical significance of the true NI_{TG} value for *GAL* gene promoters was then taken as the proportion of null groups whose promoters gave an NI_{TG} less than or equal to the true value. We assessed selection acting on non-synonymous sites in *GAL* coding sequences with a pipeline analogous to the above, except that *GAL1* and *GAL10* were considered separate loci, as the coding sequences do not overlap; thus our resampling test used 10,000 random groups of seven coding regions each. Tests for selection using the neutrality index statistic²⁴ [(polymorphic non-synonymous sites/polymorphic synonymous sites)/(divergent non-synonymous/divergent synonymous sites), with a similar formulation for non-synonymous sites] used the same

pipeline as above except that we tabulated the average neutrality index across genes of a group, and compared this average quantity for the *GAL* genes against resampled groups of the same size.

D_{XY} (ref. 25), the average number of pairwise differences between outgroup species and *S. cerevisiae* strains, and intrapopulation nucleotide diversity²⁶ (π), were calculated using custom Python scripts. Nucleotide diversity was ascertained on alignments containing only *S. cerevisiae* strains belonging to the European population, the most deeply sampled in our data set. Empirical significance of these statistics for *GAL* gene promotes and coding sequences was calculated using the resampling pipeline described above.

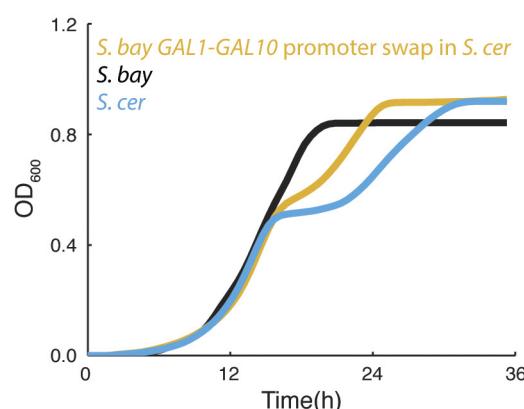
Code availability. Custom Python scripts used for data analysis are available upon request.

15. Nair, N. U. & Zhao, H. Mutagenic inverted repeat assisted genome engineering (MIRAGE). *Nucleic Acids Res.* **37**, e9 (2009).
16. Storici, F. & Resnick, M. A. The *delitto perfetto* approach to *in vivo* site-directed mutagenesis and chromosome rearrangements with synthetic oligonucleotides in yeast. *Methods Enzymol.* **409**, 329–345 (2006).
17. Warringer, J. *et al.* Trait variation in yeast is defined by population history. *PLoS Genet.* **7**, e1002111 (2011).
18. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_t}$ method. *Methods* **25**, 402–408 (2001).
19. Bradley, R. K. *et al.* Fast statistical alignment. *PLoS Comput. Biol.* **5**, e1000392 (2009).
20. Haddrill, P. R., Bachtrog, D. & Andolfatto, P. Positive and negative selection on noncoding DNA in *Drosophila* simulans. *Mol. Biol. Evol.* **25**, 1825–1834 (2008).
21. Stoletzki, N. & Eyre-Walker, A. Estimation of the neutrality index. *Mol. Biol. Evol.* **28**, 63–70 (2011).
22. Charlesworth, J. & Eyre-Walker, A. The McDonald–Kreitman test and slightly deleterious mutations. *Mol. Biol. Evol.* **25**, 1007–1015 (2008).
23. He, B. Z., Holloway, A. K., Maerkl, S. J. & Kreitman, M. Does positive selection drive transcription factor binding site turnover? A test with *Drosophila* cis-regulatory modules. *PLoS Genet.* **7**, e1002053 (2011).
24. Rand, D. M. & Kann, L. M. Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from *Drosophila*, mice, and humans. *Mol. Biol. Evol.* **13**, 735–748 (1996).
25. Nei, M. *Molecular Evolutionary Genetics* (Columbia Univ. Press, 1987).
26. Nei, M. & Li, W. H. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA* **76**, 5269–5273 (1979).

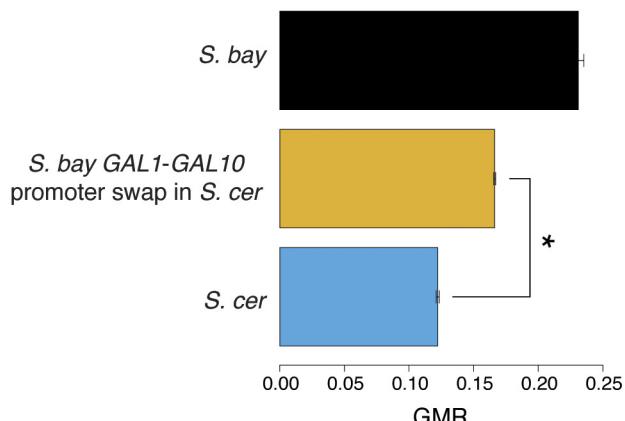


Extended Data Figure 1 | Divergence in the *S. cerevisiae* diauxic lag trait is specific to growth in glucose-galactose medium. Each trace reports growth of the indicated yeast inoculated into medium containing

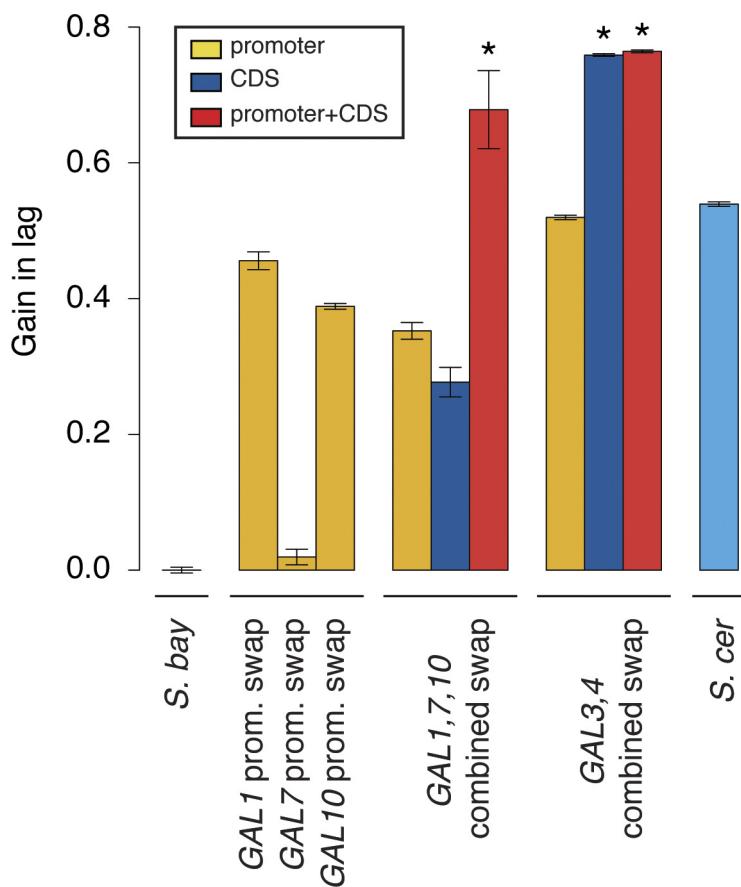
1% glucose and 1% of the indicated secondary carbon source ($n = 4$, data representative of two independent experiments).

a

Extended Data Figure 2 | The *S. bayanus* allele of the *GAL1* and *GAL10* promoters confers partial rescue of diauxic lag in *S. cerevisiae* in 1% glucose-galactose medium. **a**, Data are as in Fig. 3b, except that the yellow curve reports growth of an *S. cerevisiae* strain harbouring the *GAL1* and *GAL10* promoters from *S. bayanus* ($n \geq 12$). **b**, Each bar reports

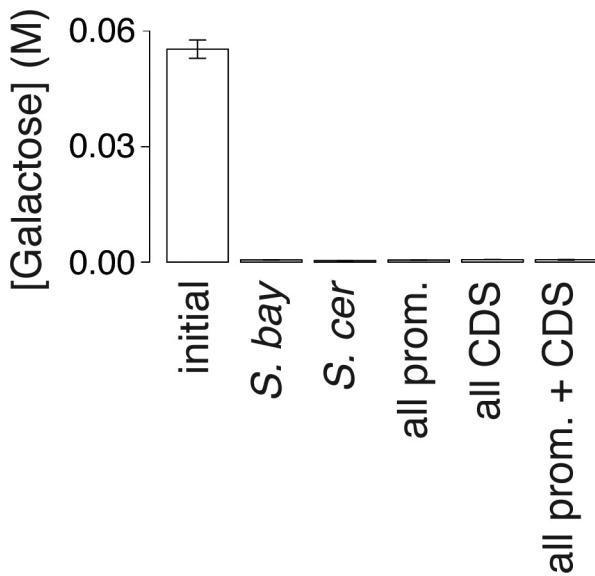
b

the GMR of the indicated strain over the time course shown in **a**. Error bars, s.e.m.; asterisk indicates a significantly different rate ($P < 1 \times 10^{-15}$, Wilcoxon rank-sum) between the transgenic swap strain and wild-type *S. cerevisiae*. Each set of data is representative of the result of two independent experiments.

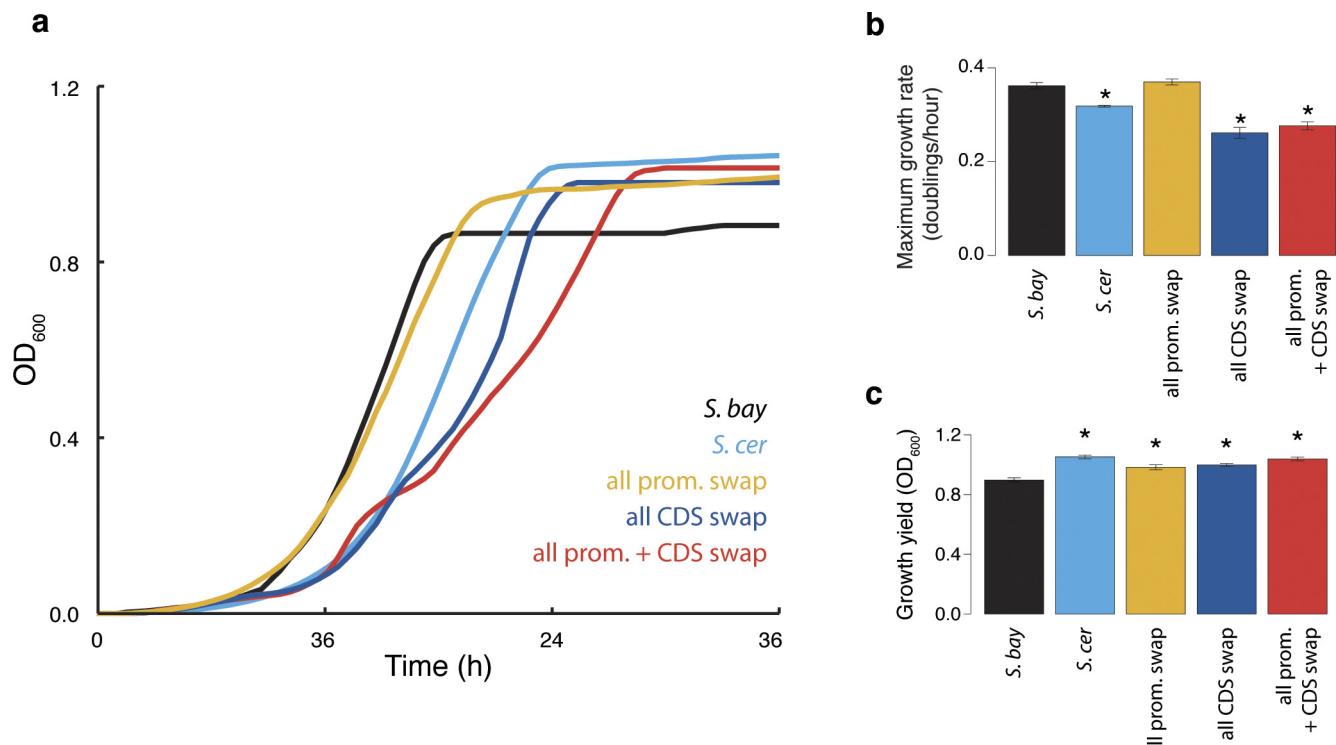


Extended Data Figure 3 | Epistasis between *S. cerevisiae* *GAL* gene alleles swapped into *S. bayanus*, in 1% glucose–galactose medium. Each bar reports the ratio of the GMR to that of wild-type *S. bayanus*, subtracted from 1, over a time course of the indicated strain inoculated into medium containing 1% glucose and 1% galactose. The first to fourth bars, and

the last bar, are from Fig. 3c; each of the remaining bars reports results from an *S. bayanus* strain harbouring *S. cerevisiae* alleles at the indicated combination of *GAL* loci ($n \geq 8$, data representative of two independent experiments). Symbols and analyses are as in Fig. 3c.



Extended Data Figure 4 | *S. bayanus* strains harbouring *S. cerevisiae* alleles of all seven *GAL* loci, inoculated into 1% glucose-galactose medium, deplete growth media of galactose. The first bar reports galactose concentration in medium before inoculation, and the remaining bars report concentrations after growth time courses, in the experiments in Fig. 3b ($n = 3$ biological replicates, each comprising 3 technical replicates).



Extended Data Figure 5 | In pure galactose medium, *S. cerevisiae* alleles of *GAL* loci are sufficient for increased biomass accumulation and, in the case of protein-coding regions, slow growth. **a**, Each trace reports growth of an *S. bayanus* strain harbouring *S. cerevisiae* alleles of all seven *GAL* genes, or a wild-type control, inoculated into medium containing 2% galactose ($n \geq 8$). **b**, Each bar reports maximum growth rate from the

respective time course in **a**. Error bars, s.e.m.; asterisks indicate values significantly different ($P < 0.05$, Wilcoxon rank-sum) from wild-type *S. bayanus*. **c**, Each bar reports growth yield from the respective time course in **a**. Error bars, s.e.m.; asterisks indicate values significantly different ($P < 0.05$, Wilcoxon rank-sum) from wild-type *S. bayanus*. Each set of data is representative of the result of two independent experiments.

Extended Data Table 1 | Excess of divergence relative to polymorphism in *GAL* promoter regions

Extended Data Table 1a

Statistic	Outgroup		
	<i>S. par</i>	<i>S. mik</i>	<i>S. bay</i>
NI _{TG} <i>GAL</i> promoters	0.32	0.37	0.31
NI _{TG} genome promoters	0.72	0.91	0.73
<i>p</i> -value	0.025	0.013	0.018

Extended Data Table 1b

Statistic	Outgroup		
	<i>S. par</i>	<i>S. mik</i>	<i>S. bay</i>
NI _{TG} <i>GAL</i> non-synonymous sites	0.98	0.92	1.03
NI _{TG} genome non-synonymous sites	1.28	1.21	1.27
<i>p</i> -value	0.25	0.19	0.26

Each panel reports analyses of the NI_{TG} measure²¹ comparing polymorphism within *S. cerevisiae* with divergence between *S. cerevisiae* and the indicated outgroup species, taken across promoter sites (a) or non-synonymous coding sites (b), with normalization by the analogous measure from synonymous coding sites. In a given panel, the first row reports NI_{TG} across the seven *GAL* genes, the second row reports the mean NI_{TG} from 10,000 randomly drawn gene groups, and the bottom row reports empirical significance of the distinction between *GAL* genes and the genomic null.

Extended Data Table 2 | Divergence between species, and polymorphism within *S. cerevisiae*, at *GAL* loci

Extended Data Table 2a

D_{XY}	Outgroup		
	<i>S. par</i>	<i>S. mik</i>	<i>S. bay</i>
<i>GAL</i> promoters	0.172	0.276	0.328
genome promoters	0.15	0.241	0.281
<i>p</i> -value	0.147	0.099	0.045
<i>GAL</i> CDS	0.102	0.155	0.188
genome CDS	0.096	0.153	0.190
<i>p</i> -value	0.337	0.429	0.573

Extended Data Table 2b

π	promoters	CDS
<i>GAL</i>	0.00067	0.00069
genome	0.00194	0.00113
<i>p</i> -value	0.181	0.513

a. The first line reports mean D_{XY} for *GAL* promoters in comparisons between *S. cerevisiae* and outgroup species. The second line reports the analogous statistics for the mean of 10,000 randomly drawn gene groups, and the third line reports empirical significance of the distinction between *GAL* genes and the genomic null. The fourth, fifth, and sixth lines are analogous to the above except that *GAL* coding regions were analysed. **b.** The first line reports π for *GAL* promoters or coding regions within *S. cerevisiae*. The second line reports the analogous statistics for the mean of 10,000 randomly drawn gene groups, and the third line reports empirical significance of the distinction between *GAL* genes and the genomic null.

Extended Data Table 3 | Evaluation of selection acting on *S. cerevisiae* promoter and coding sequences using the neutrality index statistic**Extended Data Table 3a**

Statistic	Outgroup		
	<i>S. par</i>	<i>S. mik</i>	<i>S. bay</i>
<i>GAL</i> promoters	0.32	0.42	0.35
genome promoters	0.94	1.13	0.92
<i>p</i> -value	0.003	0.009	0.014

Extended Data Table 3b

Statistic	Outgroup		
	<i>S. par</i>	<i>S. mik</i>	<i>S. bay</i>
<i>GAL</i> non-synonymous sites	1.43	1.24	1.44
genome non-synonymous sites	2.02	1.88	1.80
<i>p</i> -value	.30	.22	.36

Data are as in Extended Data Table 1 except that the neutrality index²⁴ was used for each test.

Visualization of a short-range Wnt gradient in the intestinal stem-cell niche

Henner F. Farin^{1,2,3,4}, Ingrid Jordens⁵, Mohammed H. Mosa^{2,3,4}, Onur Basak¹, Jeroen Korving¹, Daniele V. F. Tauriello⁵, Karin de Punder⁶, Stephane Angers⁷, Peter J. Peters⁶, Madelon M. Maurice^{5*} & Hans Clevers^{1*}

Mammalian Wnt proteins are believed to act as short-range signals^{1–4}, yet have not been previously visualized *in vivo*. Self-renewal, proliferation and differentiation are coordinated along a putative Wnt gradient in the intestinal crypt⁵. Wnt3 is produced specifically by Paneth cells^{6,7}. Here we have generated an epitope-tagged, functional Wnt3 knock-in allele. Wnt3 covers basolateral membranes of neighbouring stem cells. In intestinal organoids, Wnt3-transfer involves direct contact between Paneth cells and stem cells. Plasma membrane localization requires surface expression of Frizzled receptors, which in turn is regulated by the transmembrane E3 ligases Rnf43/Znrf3 and their antagonists Lgr4-5/R-spondin. By manipulating Wnt3 secretion and by arresting stem-cell proliferation, we demonstrate that Wnt3 mainly travels away from its source in a cell-bound manner through cell division, and not through diffusion. We conclude that stem-cell membranes constitute a reservoir for Wnt proteins, while Frizzled receptor turnover and ‘plasma membrane dilution’ through cell division shape the epithelial Wnt3 gradient.

Distinct mechanisms have been proposed as to how gradients of secreted Wnt protein control growth and patterning during animal development. Evidence exists for highly localized activity between adjacent cells^{1–3}, as well as for long-range activity^{8,9}. *Drosophila* Wingless (Wg) is classically considered a morphogen in the wing imaginal disk, secreted from a stripe of cells to form a long-range concentration gradient^{10–12}. This concept was recently challenged, when it was shown that membrane tethering of Wg to the producer cells did not perturb early development⁴. In vertebrates, limited information exists on endogenous Wnt protein distribution¹³, and on how Wnt gradients are built and maintained. Transcriptional Wnt target gene activity has been well-characterized in the Wnt-dependent stem-cell compartment of intestinal crypts⁵. Expression of Wnt target genes occurs in a gradient¹⁴, with highest activity at the crypt bottom where Lgr5⁺ stem cells¹⁵ are located between post-mitotic Paneth cells, that produce high levels of Wnt3 (refs 6, 7). Paneth-cell-derived Wnt3 is redundant with mesenchymally produced Wnts *in vivo*^{16,17}. In three-dimensional organoid culture, Lgr5 stem cells generate ever-growing, self-organizing ‘mini-guts’ with defined crypt- and villus-domains¹⁸, a process for which Paneth-cell-derived Wnt3 is essential¹⁶. In the current study, we exploit the mammalian mini-gut system as an experimental equivalent of the fly imaginal disk to dissect paracrine Wnt signalling.

We generated a mouse *Wnt3* allele, by introducing an haemagglutinin (HA)-tag at position Q41, located in a weakly conserved surface region opposite the Frizzled receptor binding site¹⁹ (Fig. 1a and Extended Data Fig. 1). This tagging strategy is more generally applicable to different Wnt members and peptide epitopes (Extended Data Fig. 2). While homozygous loss of *Wnt3* in mice causes gastrulation

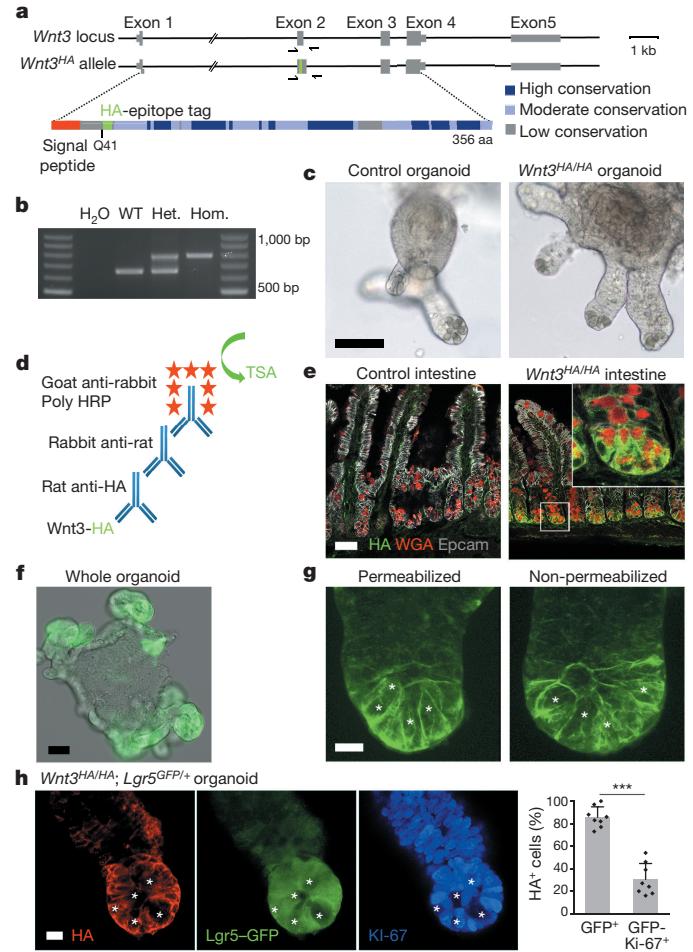


Figure 1 | Endogenous Wnt3 protein is localized to basolateral plasma membranes in the intestinal crypt. **a**, Generation of a *Wnt3*-HA knock-in allele. **b**, PCR genotyping; *Wnt3*^{HA/HA} mice are viable. WT, wild type. **c**, Morphology of organoid cultures. **d**, HA immunodetection protocol using tyramide signal amplification (TSA). **e**, Confocal imaging in intestinal tissues. Co-staining of HA, Epcam (membranes) and WGA (secretory granules). **f**, Whole-mount staining of *Wnt3*^{HA/HA} organoids. **g**, Wnt3 signal on crypt membranes is independent of permeabilization. Three-dimensional projected confocal images (see Supplementary Video 2). **h**, Staining in *Wnt3*^{HA/HA}; *Lgr5*^{EGFP+} organoids. Enriched signal on stem cells (*Lgr5*⁺-GFP⁺) compared with TA cells (*Ki-67*⁺/*Lgr5*⁺-GFP⁻). Quantification in $n=8$ organoids (mean \pm s.d.; $P < 10^{-6}$; *t*-test). Scale bars, 50 μ m (c, e, f) and 10 μ m (g, h). Paneth cells were identified on the basis of morphology and labelled by asterisks.

¹Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences (KNAW) and University Medical Center Utrecht, 3584CT Utrecht, the Netherlands. ²German Cancer Consortium (DKTK), 69120 Heidelberg, Germany. ³Georg-Speyer-Haus, Institute for Tumor Biology and Experimental Therapy, 60596 Frankfurt am Main, Germany. ⁴German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. ⁵Department of Cell Biology, Center for Molecular Medicine, University Medical Center Utrecht, 3584CX Utrecht, the Netherlands. ⁶The Maastricht Multimodal Molecular Imaging Institute, Maastricht University, 6229ER Maastricht, the Netherlands. ⁷Leslie Dan Faculty of Pharmacy, University of Toronto, Toronto, Ontario M5S 3M2, Canada.

*These authors contributed equally to this work.

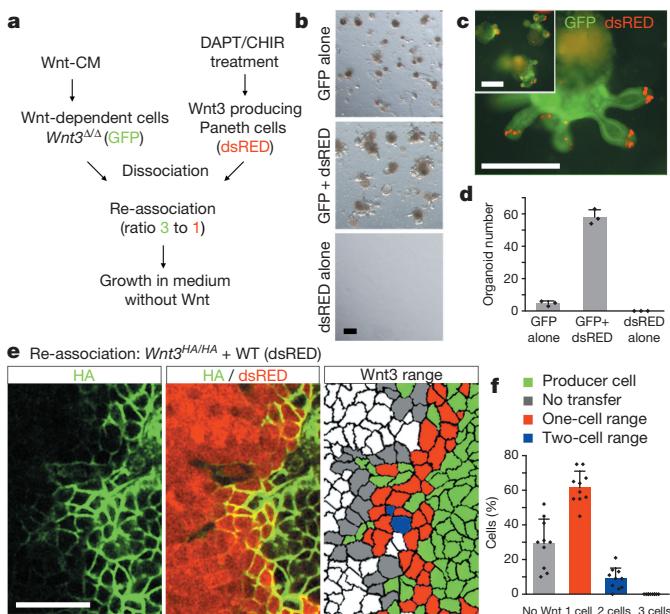


Figure 2 | Wnt3 transfer requires direct cell contact and has a limited range. **a–d**, Cell re-association of Wnt-dependent (*Wnt3*^{Δ/Δ}) with Wnt3-producing Paneth cells. **a**, Experimental protocol; **b**, light microscopic; and **c**, epifluorescent images. Self-organization of crypts with dsRED-expressing Paneth cells. **d**, Mean organoid number (\pm s.d.) in $n = 3$ independent wells after 10 days. **e**, Measurement of epithelial Wnt3-range. Re-association using Wnt3-HA producing cells (non-labelled) and wild-type cells expressing dsRED. Z-projected confocal images and colour-coding depending on the distance to the closest Wnt3-HA producing cell. **f**, Average percentage (\pm s.d.) of each distance fraction in $n = 10$ organoids. Scale bars, 200 μ m (**b**, **c**) and 25 μ m (**e**).

arrest²⁰, *Wnt3*^{HA/HA} knock-in mice are viable and fertile (Fig. 1b). To study pathway activity, we collected crypts from homozygous and control mice either directly, or after 2 weeks of mini-gut culture. Quantitative reverse transcription PCR (RT-qPCR) analysis showed no differences in the levels of Wnt target gene expression (Extended Data Fig. 3). In concordance, *Wnt3*^{HA/HA} organoids displayed normal growth and morphology (Fig. 1c). Together, the internal HA-tag fully preserved Wnt3 function.

Western blot analysis of mini-guts revealed low expression of the Wnt3-HA protein, only detectable after conversion of all cells into Wnt3-producing Paneth cells²¹ (Extended Data Fig. 4a, b). On intestinal sections, a multi-step amplification protocol allowed specific Wnt3-HA detection in Paneth cells (Fig. 1d, e). Paneth cell granules were negative, implying that Wnt3 is not secreted via this dominant, apical secretion route (Extended Data Fig. 4c, d and Supplementary Video 1). To obtain high-resolution information in a system amenable to experimental manipulation, we derived mini-guts from homozygous knock-in mice. We observed fluorescence exclusively in the crypt equivalents within the mini-guts, while the villus-domain was negative (Fig. 1f). Confocal laser microscopy demonstrated that Wnt3 covers basolateral crypt membranes (Fig. 1g, Extended Data Fig. 4e, f and Supplementary Video 2). This staining pattern was not dependent on membrane permeabilization, indicating that Wnt3 is mainly localized on the external cell surface. Co-localization with the *Lgr5*^{EGFP} reporter¹⁵ and KI-67 showed a specific enrichment on *Lgr5* stem cells compared with transit-amplifying (TA) cells (Fig. 1h).

To determine whether the mini-gut cultures secrete diffusible Wnt activity, we tested if Wnt3-producing, wild-type organoids could rescue the growth of *Wnt3*^{Δ/Δ} (knockout) cultures. Fluorescently labelled mini-guts were co-seeded in Matrigel. We found that the presence of wild type (dsRED) did not support growth of *Wnt3*^{Δ/Δ} mini-guts (green fluorescent protein, GFP) (Extended Data Fig. 5a–d), arguing that diffusible Wnt activity was negligible. Consistently, we could

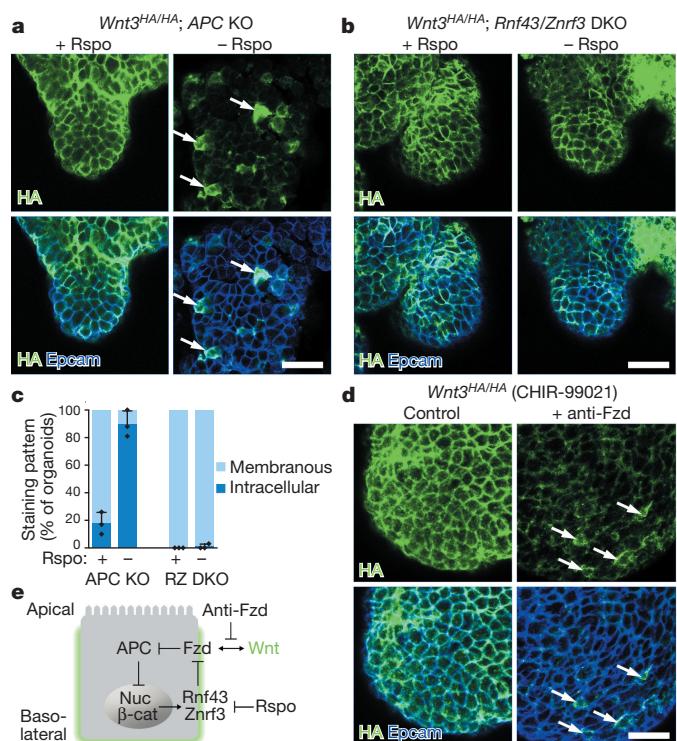


Figure 3 | Frizzled receptors act as membrane tether of Wnt. **a, b**, R-spondin (Rspo) controls membrane localization in APC knockout (**a**) but not in *Rnf43/Znrf3* double mutants (**RZ DKO**; **b**). Epcam staining of cell membranes; arrows show intracellular staining. **c**, Mean percentage of organoids (\pm s.d.) with membranous or intracellular staining. Data from $n = 3$ independent experiments. **d**, Neutralizing pan-Frizzled antibody (anti-Fzd) reduces surface expression of Wnt3-HA with residual intracellular signal (arrows). Treatment for 48 h; CHIR-99021 was added to rescue organoid growth. The z-projections of confocal images are shown; all scale bars, 25 μ m. **e**, Model for control of the Wnt3 localization by Frizzled and the E3 ligases *Rnf43/Znrf3*.

neither detect transfer of Wnt3-HA between organoids, nor interfere with growth of *Wnt3*^{HA/HA} cultures by embedding anti-HA affinity beads that efficiently sequester diffusible factor (Extended Data Fig. 5e, f).

To investigate direct protein transfer between adjacent epithelial cells, we performed re-association assays after single-cell dispersal of fluorescently labelled organoids. GFP-expressing *Wnt3*^{Δ/Δ} cells were mixed with *in vitro* differentiated wild-type Paneth cells (dsRED). Aggregates were transferred to Matrigel and cultured in the absence of exogenous Wnt (see experimental scheme in Fig. 2a). Wild-type Paneth cells rescued the growth of *Wnt3*^{Δ/Δ} cells (Fig. 2b–d). We observed efficient self-organization into crypt structures with a stereotypical arrangement of dsRED-labelled cells at the budding tips (Fig. 2c), consistent with the model that Paneth cells serve as symmetry-breaking crypt organizers in culture. Staining revealed spatial confinement of Wnt3-HA around individual Paneth cells (Supplementary Video 3). To measure the extent of Wnt3 propagation within the epithelial sheet more easily, we generated spherical organoids (by addition of Wnt3a-conditioned medium (CM))⁷. Interfaces between Wnt3-HA-producing cells (non-labelled) and wild-type receiving cells (dsRED-labelled, which only express non-tagged Wnt) were readily recognizable. Wnt3-HA penetrated one cell diameter (two at the most) into the Wnt3-HA-negative domains (Fig. 2e, f). Strong differences in Wnt3-HA decoration were observed between adjacent cells. Thus, localized production and limited epithelial spreading appeared important hallmarks of Wnt3 distribution.

We subsequently investigated if Wnt3-HA binding to stem-cell membranes depends on their cognate Frizzled receptors. As crypts

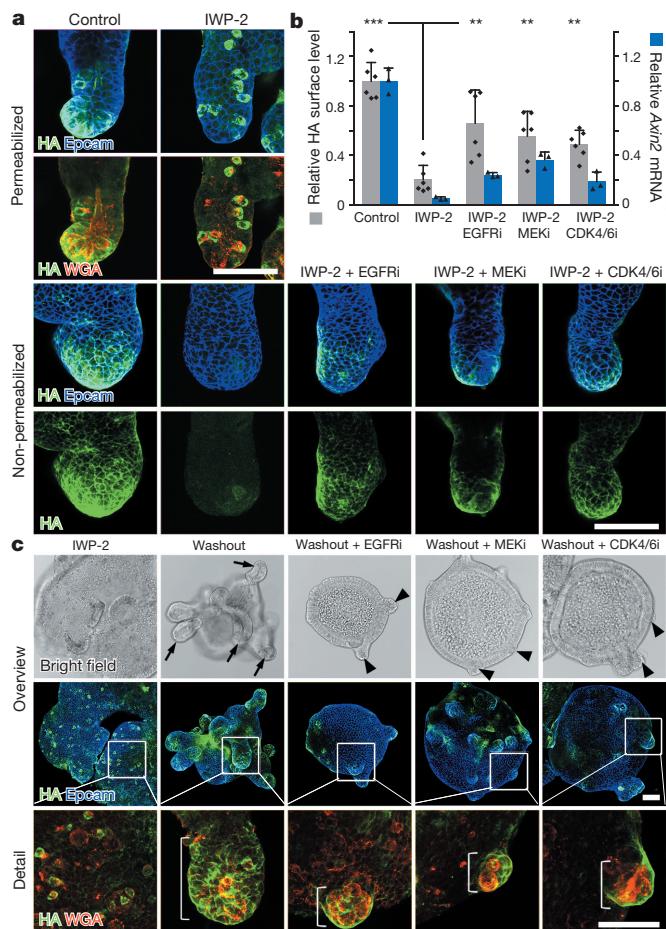


Figure 4 | Cell proliferation influences Wnt3 surface level and signal range. **a**, Chase experiment: 24 h after block of Wnt-production (IWP-2), Wnt3-HA is sequestered in WGA-positive Paneth cells (top) and absent from the cell surface (non-permeabilized samples; bottom). Parallel block of proliferation by inhibitors of EGFR, MEK or CDK4/6 causes retention of surface HA signal. **b**, Quantification of Wnt3 surface level and transcriptional activity. Grey bars, mean HA-staining intensity in $n = 6$ non-permeabilized organoids per condition. Blue bars, mean *Axin2* expression in $n = 3$ independent wells; error bars, s.d.; *t*-test compared with IWP-2 condition: *** $P < 10^{-3}$; ** $P < 10^{-2}$. **c**, Pulse experiment after Wnt-release (washout of IWP-2/Wnt-CM). Crypts formed in normal medium (arrows) are positive for Wnt3-HA. After block of proliferation, rudimentary crypts are formed (arrowheads) and the Wnt3-HA signal remains focused (brackets). Z-projected confocal images. All scale bars, 50 μ m.

express multiple Frizzleds⁶, we altered their surface expression indirectly by modulation of the transmembrane E3-ligases Rnf43 and Znrf3. These act as negative feedback regulators of Wnt signalling by ubiquitylation and subsequent lysosomal degradation of Frizzleds^{22,23}. Of note, Lgr4/5 proteins and their R-spondin ligands antagonize Frizzled downregulation by recruiting Rnf43/Znrf3 into a trimeric complex^{22,24}. Given that organoid growth requires R-spondin¹⁸, we first tested its effect on Wnt localization. We found that the Wnt3-HA signal expanded or decreased in an R-spondin1 concentration-dependent manner (Extended Data Fig. 6a). Because *Wnt3* transcription was also strongly R-spondin-dependent (Extended Data Fig. 6b), we aimed to uncouple the effect of R-spondin on Frizzled-turnover from downstream signalling. To this end, we genetically activated the pathway in *Wnt3*^{HA/HA} cultures. CRISPR/Cas9-induced null mutations in *APC* or in *Rnf43/Znrf3* (Extended Data Fig. 7)²⁵ resulted in R-spondin-independent growth. Wnt3-HA was absent from the surface of *APC* mutant mini-guts (that express high levels of Rnf43/Znrf3), but could be restored by addition of

R-spondin (Fig. 3a, c). This implied that Frizzleds tether Wnt3 to membranes, as has been suggested in the *Drosophila* wing disc²⁶. Consistently, Wnt3 surface localization was high and insensitive to R-spondin in *Rnf43/Znrf3* mutant cultures (Fig. 3b, c). In cell re-association assays, Wnt3-HA transfer to *Rnf43/Znrf3* mutant cells was increased and to *APC* mutant cells was reduced (Extended Data Fig. 8). Independently, the application of a neutralizing pan-Frizzled antibody²⁷ strongly reduced Wnt3-HA surface expression (Fig. 3d). Together, these experiments indicate that surface-expressed Frizzled binds and retains Wnt3 (Fig. 3e).

To explore the kinetics of Wnt3 surface expression, we acutely blocked Wnt secretion using the Porcupine inhibitor IWP-2 (ref. 28). After 24 h, Wnt3-HA was observed exclusively inside Paneth cells, highlighting the requirement of Porcupine for Wnt trafficking to the plasma membrane of these cells (Fig. 4a). Staining of non-permeabilized organoids confirmed that surface-bound Wnt3 was absent, indicating that its half-life is below 24 h. To test if cell division has an impact on propagation of surface-bound Wnt3, we induced cell cycle arrest using the EGFR-receptor inhibitor Gefitinib, the downstream MEK inhibitor PD-032501, or the CDK4/6 inhibitor Palbociclib (Extended Data Fig. 9a, b). In all cases, residual Wnt3 surface signal and increased transcriptional activity was observed 24 h after initiation of IWP-2 treatment (Fig. 4a, b), indicating that cell division under normal conditions dilutes cell-bound Wnt3.

To address if cell division is indeed the means by which Wnt3 is propagated from the Paneth cell source, we blocked Wnt secretion for 3 days, followed by washout of the Porcupine inhibitor. Formation of multiple crypt-like structures, invariably decorated with a gradient of Wnt3-HA, was observed within 2 days after IWP-2 washout (Fig. 4c and Supplementary Video 4). In the presence of EGFR, MEK or CDK4/6 inhibitors, crypt formation was blocked and Wnt3-HA remained localized around the producing cells. These results suggest that signal propagation is dependent on cell division.

From these combined observations, we concluded that two phenomena control the size and shape of the short-range, graded Wnt3 signal. First, Wnt3, produced by Paneth cells, does not freely diffuse, but is transferred to the nearest neighbour, typically an Lgr5 stem cell. Frizzled receptors tether Wnt to the membrane of the receiving stem cell. Surface expression of the Wnt-Frizzled complex is negatively controlled by Rnf43/Znrf3, which in turn can be alleviated by Lgr4/5 and R-spondin. Second, Frizzled-bound Wnt spreads passively by stem-cell division, which dilutes surface-bound Wnt and thus creates a gradient.

Similar highly localized Wnt signals have been described in invertebrate models^{1–3}. Long-range signalling is dispensable during fly embryogenesis, but transfer over longer distances becomes important during post-juvenile organ growth⁴. Apparently, the body dimensions then become too large to support Wnt-signalling only by auto/juxtacrine sources. From an evolutionary perspective it is interesting that the R-spondin/Lgr Wnt signal amplifier is a vertebrate-specific innovation, while Rnf43 is already present in *Caenorhabditis elegans*. Organ size in vertebrates may require amplified cellular output from stem-cell niche compartments. The R-spondin/Lgr signalling module appears designed to prolong/increase Wnt surface levels once cells have left contact to the niche to build a TA compartment. Unlike the *Drosophila* wing disc, where autocrine Wnt signalling appears to predominate⁴, Wnt signalling is non-cell-autonomous and relies on the presence of dedicated niche Paneth cells as well as non-epithelial sources^{16,17}. Transfer within the epithelium requires direct Paneth cell contact (this study), a notion that is in agreement with our previous observation that stem cells attempt to maximize their membrane contact to Paneth cells⁷. This also explains why a central position of Lgr5 stem cells between Paneth cells supports stem-cell maintenance more efficiently than a peripheral position²⁹. It has been demonstrated that topical application of Wnt (immobilized on a bead) to one side of a stem cell results in an asymmetrical outcome of the subsequent stem-cell division³⁰. Crypt stem cells located at the niche boundary

experience a similarly sharp Wnt gradient, as only one side of a 'boundary stem cell' touches its Wnt source, the Paneth cell. Stem-cell plasma membranes thus carry positional information that could be involved in timing of differentiation. How Wnt3 is transferred between producing and receiving cells remains to be determined.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 October; accepted 18 December 2015.

Published online 10 February 2016.

- van den Heuvel, M., Nusse, R., Johnston, P. & Lawrence, P. A. Distribution of the wingless gene product in *Drosophila* embryos: a protein involved in cell-cell communication. *Cell* **59**, 739–749 (1989).
- Thorpe, C. J., Schlesinger, A., Carter, J. C. & Bowerman, B. Wnt signaling polarizes an early *C. elegans* blastomere to distinguish endoderm from mesoderm. *Cell* **90**, 695–705 (1997).
- Goldstein, B., Takeshita, H., Mizumoto, K. & Sawa, H. Wnt signals can function as positional cues in establishing cell polarity. *Dev. Cell* **10**, 391–396 (2006).
- Alexandre, C., Baena-Lopez, A. & Vincent, J.-P. Patterning and growth control by membrane-tethered Wingless. *Nature* **505**, 180–185 (2014).
- Clevers, H. The intestinal crypt, a prototype stem cell compartment. *Cell* **154**, 274–284 (2013).
- Gregorieff, A. et al. Expression pattern of Wnt signaling components in the adult intestine. *Gastroenterology* **129**, 626–638 (2005).
- Sato, T. et al. Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. *Nature* **469**, 415–418 (2011).
- Coudreuse, D. Y. M., Roël, G., Betist, M. C., Destré, O. & Korswagen, H. C. Wnt gradient formation requires retromer function in Wnt-producing cells. *Science* **312**, 921–924 (2006).
- Kiecker, C. & Niehrs, C. A morphogen gradient of Wnt/β-catenin signalling regulates anteroposterior neural patterning in *Xenopus*. *Development* **128**, 4189–4201 (2001).
- Zecca, M., Basler, K. & Struhl, G. Direct and long-range action of a wingless morphogen gradient. *Cell* **87**, 833–844 (1996).
- Neumann, C. J. & Cohen, S. M. Long-range action of Wingless organizes the dorsal-ventral axis of the *Drosophila* wing. *Development* **124**, 871–880 (1997).
- Strigini, M. & Cohen, S. M. Wingless gradient formation in the *Drosophila* wing. *Curr. Biol.* **10**, 293–300 (2000).
- Galli, L. M., Barnes, T. L., Secrest, S. S., Kadowaki, T. & Burrus, L. W. Porcupine-mediated lipid-modification regulates the activity and distribution of Wnt proteins in the chick neural tube. *Development* **134**, 3339–3348 (2007).
- Muñoz, J. et al. The Lgr5 intestinal stem cell signature: robust expression of proposed quiescent '+4' cell markers. *EMBO J.* **31**, 3079–3091 (2012).
- Barker, N. et al. Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* **449**, 1003–1007 (2007).
- Farin, H. F., Van Es, J. H. & Clevers, H. Redundant sources of Wnt regulate intestinal stem cells and promote formation of Paneth cells. *Gastroenterology* **143**, 1518–1529.e7 (2012).
- Kabiri, Z. et al. Stroma provides an intestinal stem cell niche in the absence of epithelial Wnts. *Development* **141**, 2206–2215 (2014).
- Sato, T. et al. Single Lgr5 stem cells build crypt-villus structures *in vitro* without a mesenchymal niche. *Nature* **459**, 262–265 (2009).
- Janda, C. Y., Waghray, D., Levin, A. M., Thomas, C. & Garcia, K. C. Structural basis of Wnt recognition by Frizzled. *Science* **337**, 59–64 (2012).
- Liu, P. et al. Requirement for Wnt3 in vertebrate axis formation. *Nature Genet.* **22**, 361–365 (1999).
- Yin, X. et al. Niche-independent high-purity cultures of Lgr5⁺ intestinal stem cells and their progeny. *Nature Methods* **11**, 106–112 (2014).
- Hao, H.-X. et al. ZNRF3 promotes Wnt receptor turnover in an R-spondin-sensitive manner. *Nature* **485**, 195–200 (2012).
- Koo, B.-K. et al. Tumour suppressor RNF43 is a stem-cell E3 ligase that induces endocytosis of Wnt receptors. *Nature* **488**, 665–669 (2012).
- Chen, P.-H., Chen, X., Lin, Z., Fang, D. & He, X. The structural basis of R-spondin recognition by LGR5 and RNF43. *Genes Dev.* **27**, 1345–1350 (2013).
- Schwank, G. et al. Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. *Cell Stem Cell* **13**, 653–658 (2013).
- Lecourtois, M., Alexandre, C., Dubois, L. & Vincent, J.-P. Wingless capture by Frizzled and Frizzled2 in *Drosophila* embryos. *Dev. Biol.* **235**, 467–475 (2001).
- Gurney, A. et al. Wnt pathway inhibition via the targeting of Frizzled receptors results in decreased growth and tumorigenicity of human tumors. *Proc. Natl. Acad. Sci. USA* **109**, 11717–11722 (2012).
- Chen, B. et al. Small molecule-mediated disruption of Wnt-dependent signaling in tissue regeneration and cancer. *Nature Chem. Biol.* **5**, 100–107 (2009).
- Ritsma, L. et al. Intestinal crypt homeostasis revealed at single-stem-cell level by *in vivo* live imaging. *Nature* **507**, 362–365 (2014).
- Habib, S. J. et al. A localized Wnt signal orients asymmetric stem cell division *in vitro*. *Science* **339**, 1445–1448 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank M. van den Born, C. Kroon-Veenboer, J. van Es, T. Darvishi and S. van den Brink and the Hubrecht imaging facility for technical support. We thank A. Gurney for the pan-Fzd antibody and B.-K. Koo for discussions. M.M.M. is supported by the European Research Council (ERC StG 242958), the European Union (FP7 Marie-Curie ITN 608180 'WntsApp') and the Netherlands Scientific Organization (NWO-ECHO 711.013.012 and NWO-VICI 91815604).

Author Contributions H.F.F. conceived the study, performed the experiments and wrote the manuscript. I.J. and O.B. performed experiments and provided intellectual input. M.H.M. and D.V.F.T. performed experiments. J.K. performed ES blastocyst injection. K.P., S.A. and P.J.P. provided microscopic support and reagents. M.M.M. and H.C. conceived the study and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.C. (h.clevers@hubrecht.eu) or H.F.F. (farin@gsh.uni-frankfurt.de).

METHODS

Mice. For generation of the *Wnt3*^{HA} allele, a 1,044-base-pair genomic fragment covering the exon 2 of the gene using the primer pair forward 5'-TCCTTTGTCCTTATTTGGATTC-3' and reverse 5'-AGGGAAATGTCA CACATGTCTAC-3' was subcloned and the HA-tag was introduced using the following PCR primers: forward 5'-GGCACGTCGTATGGGTA CTGGGAGGCCAGAGATGTGTC-3' and reverse 5'-AGACTACGGCATCTCT GCCTCTGCTCTGGGCTCCATC-3' (underlined residues encoding for the epitope tag). The fragment was subsequently amplified with primers flanked by LoxP and LoxM sites, and introduced in reverse orientation into the pL451 plasmid that contained a Pgk-neomycin-pA selection cassette flanked by FRT-sites by recombination cloning (In-Fusion HD, Clontech). The wild-type exon 2 and the 5'- and 3'-homology regions were introduced in a configuration that allowed excision of the wild-type exon 2 and inversion of the HA-tagged exon 2 after Cre-mediated recombination between the two pairs of LoxP and LoxM sites (Extended Data Fig. 1c). After sequence verification the linearized targeting vector was used for homologous recombination in embryonic stem cells that was confirmed by Southern blot analysis (Extended Data Fig. 1d). After blastocyst injection, chimaeric offspring was crossed to FLPeR mice³¹ to remove the neomycin expression cassette. Germline transmission was confirmed by PCR and mice were crossed to *Pgk-Cre*³² mice to induce germline recombination of the *Wnt3*^{HA} allele. The *Lgr5*^{EGFP-IRES-CreERT2} allele, the conditional *Wnt3*^{fl} allele and the transgenic *Vil-CreERT2* line have been described^{15,33,34}. Alleles were maintained on a mixed C57BL/6 background and littermate groups of mixed sexes were analysed at ages between 8 and 12 weeks. Experiments were performed according to guidelines and reviewed by the Animal Experiments Committee (DEC) of the Royal Netherlands Academy of Arts and Sciences.

Organoid culture. Mouse organoids were established and maintained as described^{16,18} from isolated crypts collected from the entire length of the small intestine. The basic culture medium (ENR) contained advanced DMEM/F12 supplemented with penicillin/streptomycin, 10 mM HEPES, 1× Glutamax, 1× B27 (all from Life Technologies) and 1 mM N-acetylcysteine (Sigma) that was supplemented with murine recombinant EGF (Peprotech), R-spondin1-CM (5% final volume if not indicated otherwise) and Noggin-CM (10% v/v). *Wnt3a*-CM was used where specified at 50% (v/v), if not indicated otherwise. A mycoplasma-free status was confirmed routinely. *Wnt3*-null cultures were obtained from *Vil-CreERT2*; *Wnt3*^{fl/fl} organoids after treatment with 0.5 μM 4-OH-tamoxifen overnight and were then maintained by addition of *Wnt3a*-CM. For constitutive lentiviral expression of mouse *Wnt3*, *Wnt3*-HA and *Wnt3*-Flag, the open reading frames were inserted 3' of a *Pgk* promoter in the lentiviral vector pLVIRES-puro. For cell labelling *Pgk*:EGFP-IRES-puro or *Pgk*:dsRED-IRES-puro containing lentivirus was used and transduced as described³⁵. Puromycin (Invivogen) was included in the medium (0.5–1 μg ml⁻¹) and fluorescent reporter expression was monitored and documented using an EVOS fl inverse fluorescence microscope (Life Technologies).

For bead neutralization experiments, the affinity matrix (anti-HA (clone 3F10)-coupled agarose beads, Roche) was washed thoroughly in medium before resuspension in Matrigel (in a ratio 1:6 (v/v)) and organoid embedding. Culture media were pre-incubated with 50 μl ml⁻¹ washed beads before addition. For cell re-association experiments, cells were collected 5 days after seeding, either in regular medium containing *Wnt*-CM or Paneth cell differentiation medium²¹ in the presence of 1 μM CHIR-99021 (Stemgent) and 15 μM DAPT (Sigma-Aldrich) in TrypLE Express (Life Technologies) and digested for 2 × 5 min at 37 °C followed by gentle mechanical dispersal. Cells were washed in cold medium containing 5% FBS and resuspended in medium containing 1,000 U ml⁻¹ DNase I (Roche) before filtration and centrifugation. The cells were then resuspended in medium containing 10% dissolved Matrigel (BD), 1 μM CHIR-99021 and 10 μM Y-27632 (Sigma-Aldrich). Eight thousand Paneth cells and/or 24,000 GFP-positive cells were co-seeded in 48-well suspension plates that had been pre-blocked with 5% FBS/PBS. After 24 h of re-association, the aggregates were collected, seeded in Matrigel and cultured in the presence of 1 μM CHIR-99021 which was withdrawn after 48 h. Organoid number was counted after 10 days and the experiment was replicated twice. For staining of re-associated organoids 25% (v/v) *Wnt*-CM was added to the normal culture medium to generate large mosaic spheroids, containing increased number of Paneth cells¹⁶. Here dsRED-labelled cells and non-labelled *Wnt3*^{HA/HA} cells were mixed in a 2:1 ratio. Blocking pan-Frizzled antibody (clone OMP-18R5 (ref. 27)) was added to the culture medium at a concentration of 50 μg ml⁻¹ in the presence of 3 μM CHIR-99021. For *Wnt3*-HA chase experiments, 2.5 μM IWP-2 (Stemgent) was added to the normal medium. Gefitinib (AstraZeneca, 1 μM), PD-0325901 (Sigma, 2 μM) and Palbociclib (PD 0332991, Sigma, 1.5 μM) were added 6 h before IWP-2-treatment to block the cell cycle. For pulse experiments, organoids were cultured in medium containing

IWP-2 for 3 days before washing and growth in normal culture medium with cell cycle inhibitors (as above). Time-lapse videos were recorded on a Leica AF7000 microscope using 30 min intervals.

CRISPR/Cas9 mutagenesis. Mutations in mouse organoids using the CRISPR/Cas9 technology were induced as described²⁵. For targeting of mouse *APC*, the reported single guide RNA (sgRNA)-5 and for *Rnf43* and *Znrf3* sgRNA-3 and sgRNA-2 were used, respectively. Correct targeting was confirmed by sequencing of *Escherichia coli*-cloned genomic PCR products (Extended Data Fig. 7). For each genotype, two independent clones were studied.

Immunostaining. Intestinal tissues were fixed 4% formaldehyde/PBS for 20 min at room temperature (22 °C) followed by embedding Tissue-Tek O.C.T. compound. Cryosections (50 μm) were incubated with Mouse on Mouse Blocking Reagent (Vector Laboratories) and then with rat anti-HA monoclonal antibody (clone 3F10; Roche; 0.2 μg ml⁻¹) overnight at 4 °C, followed by incubation with unconjugated rabbit anti-rat IgG(H+L) (SouthernBiotech, 2.5 μg ml⁻¹, 1 h at room temperature) and BrightVision Poly-HRP-Anti Rabbit reagent (ImmunoLogic, 1 h at room temperature) that were both pre-blocked using normal mouse serum. For detection TSA (Life Technologies) was used according to the supplier's instructions with Alexa Fluor 488 tyramide diluted 1:100 in amplification buffer. Wheat Germ Agglutinin, Texas Red-X Conjugate (Life Technologies, 5 μg ml⁻¹) was applied for 1 h at room temperature before the slides were embedded using ProLong Gold Antifade reagent (Life Technologies).

For whole-mount stainings, organoids were collected in ice-cold medium, pelleted and resuspended in chilled Cell Recovery Solution (BD). After incubation (15 min on ice) and gentle mixing, organoids were washed in medium, pelleted, resuspended and fixed in 2% formaldehyde/PBS overnight at 4 °C. Organoids were subsequently permeabilized in PBS 0.1% Tween 20 (30 min at room temperature) before resuspension in blocking solution containing 0.2% normal donkey serum/PBS at 4 °C for 1 h. After incubation in anti-HA antibody (as above, 0.5 μg ml⁻¹, 4 °C overnight), three wash steps were performed at room temperature by addition of ice-cold PBS 0.1% Tween 20 and organoid sedimentation. For staining without permeabilization, organoids were fixed overnight in 2% formaldehyde/PBS containing 4% sucrose before washes in blocking buffer and incubation with primary antibody in the absence of detergent. Subsequent incubation steps for both protocols were rabbit anti-rat IgG(H+L) (as above) and BrightVision Poly-HRP-Anti Rabbit reagent (as above; 1:2 diluted in PBS with final 0.1% Tween 20, 0.1% Triton X-100). At this step Wheat Germ Agglutinin (as above), 4',6-diamidino-2-phenylindole (DAPI), eFluor-660-conjugated rat anti-Ki67 (eBioscience; clone SolA15; 0.4 μg ml⁻¹) and/or anti-mouse-Epcam-APC (eBioscience; clone G8.8; 0.5 μg ml⁻¹) were added, followed by washes and pre-equilibration in TSA amplification buffer (PBS pH 7.6 with 0.1 mM imidazole). Specimens were incubated for 30 min at room temperature in a 1:1,500 dilution of TSA (as above or using TSA-Cy3; PerkinElmer), followed by washes and mounting (as above). Specimens were documented on a Leica SP8 scanning confocal microscope using a ×20 objective and z-step size of 1 μm. A ×63 objective with 0.5 μm z-steps was used for Extended Data Fig. 4d, f.

Image analysis. Fiji software was used to generate z- and three-dimensional projections that were exported as video files. Channels were overlaid using Adobe Photoshop CS6. For co-localization analysis, presence of *Lgr5*-GFP, anti-HA and anti-Ki-67 staining was determined in each DAPI-positive cell on confocal z-projections of crypt hemispheres ($n=8$ representative specimens). The *Wnt3*-HA range in mosaic organoids was determined on z-projected confocal images. For this, Epcam signals were imported into CellProfiler software³⁶ and analysed using the example pipeline 'Tissue Neighbours' that was adapted to trace all cell outlines. This mask was overlaid on dsRED and Alexa Fluor 488 signals to identify *Wnt3*-HA producing (dsRED negative) and *Wnt3*-HA receiving (dsRED positive) cells. Alexa Fluor 488 signals were thresholded to determine the cell-distance to the nearest producing cell: receiving cells were labelled positive only if a signal was found on a membrane side not in direct contact with an adjacent *Wnt*-HA producing cell. To omit cells that were engulfed by producers, receiving cells were excluded if they had four or more dsRED-negative direct neighbours. Cells were counted as 'no transfer' events only if they had direct dsRED-negative neighbour(s). The distance fractions were determined in each $n=10$ representative mosaic organoids from experiments that were repeated three times. For surface-level measurement, *Wnt3*-HA and Epcam images were recorded using constant microscope settings. Pixel intensities (whole image) were measured using Fiji. For each specimen the *Wnt3*-HA intensity was normalized to the Epcam intensity to adjust for organoid size. Six representative crypt hemispheres were recorded per condition and the experiments were repeated three times.

Western blotting, *Wnt*-conditioned media and reporter assay. HEK293T cells were transfected with pcDNA3 expression vector that encoded the ORF of the mouse *Wnt3* complementary DNA (cDNA), which contained an HA epitope at the corresponding position (Q41) that was introduced by PCR mutagenesis with

the primers described above. Whole cell/organoid lysates were collected in 1× denaturating SDSPage Buffer, before western blot analysis using anti-HA (as above, 0.5 µg ml⁻¹ in PBS containing 5% milk powder). For generation of Flag-tagged (Extended Data Fig. 2c) or HA-tagged (Extended Data Fig. 5f) mouse Wnt3a, an internal epitope was inserted at the equivalent position (Q38). A construct with HA-tagged mouse Wnt3a was used for Extended Data Fig. 5f. For production of Wnt-conditioned media, murine L-cells were transfected with Wnt3a constructs using FuGENE (Promega). Stable, clonal cell lines were established after selection using Zeocine. Media were harvested after 3–4 days of conditioning and analysed by western blotting using mouse anti-Wnt3a (a gift from R. Takada) and mouse anti-Flag M2 (Sigma) antibodies. HEK293T cells were transiently transfected with TOPFlash or FOPFlash reporter constructs and TK-Renilla as a transfection control. Cells were stimulated overnight with either control L-cell medium or Wnt3a-conditioned medium. Luciferase activity was measured using the Dual Luciferase Reporter kit (Promega) according to the manufacturer's protocol.

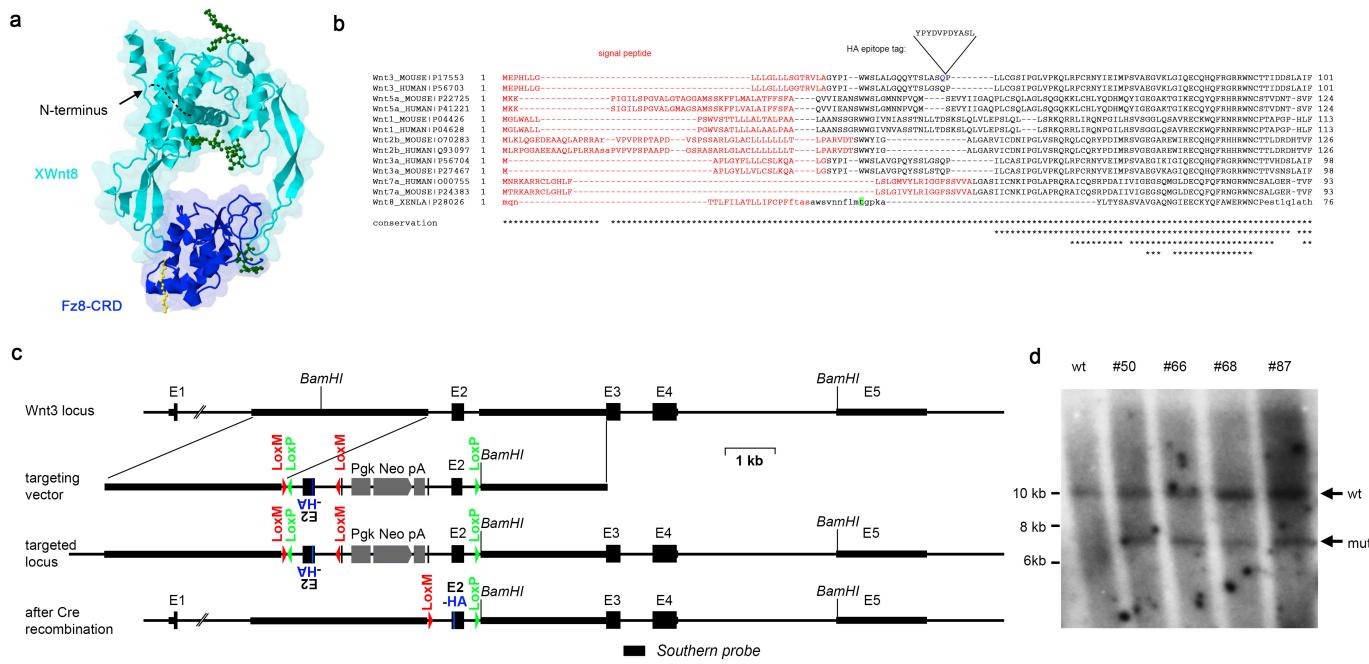
qPCR analysis. RNA preparation, cDNA preparation, qPCR and primer sequences have been described¹⁶.

EDU incorporation assay. EDU incorporation assay was performed using the Click-iT EdU Alexa Fluor 488 Flow Cytometry Assay Kit (Life Technologies). Cells were treated with 10 µM EDU 1 h before harvesting and single cell dispersal (as above) for fluorescence-activated cell sorting (FACS) analysis. Genomic DNA content was measured by addition 0.5 µg ml⁻¹ 7-AAD (eBioscience) before analysis.

Statistical analysis. Sample size was chosen empirically following previous experience in the assessment of experimental variability. No statistical methods were used to predetermine sample size. Samples were not randomized and the

investigators were not blinded. After the normal distribution was confirmed using the Shapiro–Wilk test, significant differences between two groups were evaluated using two-tailed, unpaired Student's *t*-tests. No samples/specimens were excluded from the statistical analysis. Differences were considered to be significant when *P* < 0.05.

31. Farley, F. W., Soriano, P., Steffen, L. S. & Dymecki, S. M. Widespread recombinase expression using FLPeR (flipper) mice. *Genesis* **28**, 106–110 (2000).
32. Lallemand, Y., Luria, V., Haffner-Krausz, R. & Lonai, P. Maternally expressed PGK-Cre transgene as a tool for early and uniform activation of the Cre site-specific recombinase. *Transgenic Res.* **7**, 105–112 (1998).
33. Barrow, J. R. et al. Ectodermal Wnt3/β-catenin signaling is required for the establishment and maintenance of the apical ectodermal ridge. *Genes Dev.* **17**, 394–409 (2003).
34. El Marjou, F. et al. Tissue-specific and inducible Cre-mediated recombination in the gut epithelium. *Genesis* **39**, 186–193 (2004).
35. Koo, B.-K. et al. Controlled gene expression in primary Lgr5 organoid cultures. *Nature Methods* **9**, 81–83 (2012).
36. Lamprecht, M. R., Sabatini, D. M. & Carpenter, A. E. CellProfiler: free, versatile software for automated biological image analysis. *Biotechniques* **42**, 71–75 (2007).
37. Jmol: an open-source Java viewer for chemical structures in 3D. <http://www.jmol.org/>
38. Morgenstern, B. DIALIGN: multiple DNA and protein sequence alignment at BiBiServ. *Nucleic Acids Res.* **32**, W33–W36 (2004).
39. Schepers, A. G., Vries, R., van den Born, M., van de Wetering, M. & Clevers, H. Lgr5 intestinal stem cells have high telomerase activity and randomly segregate their chromosomes. *EMBO J.* **30**, 1104–1109 (2011).

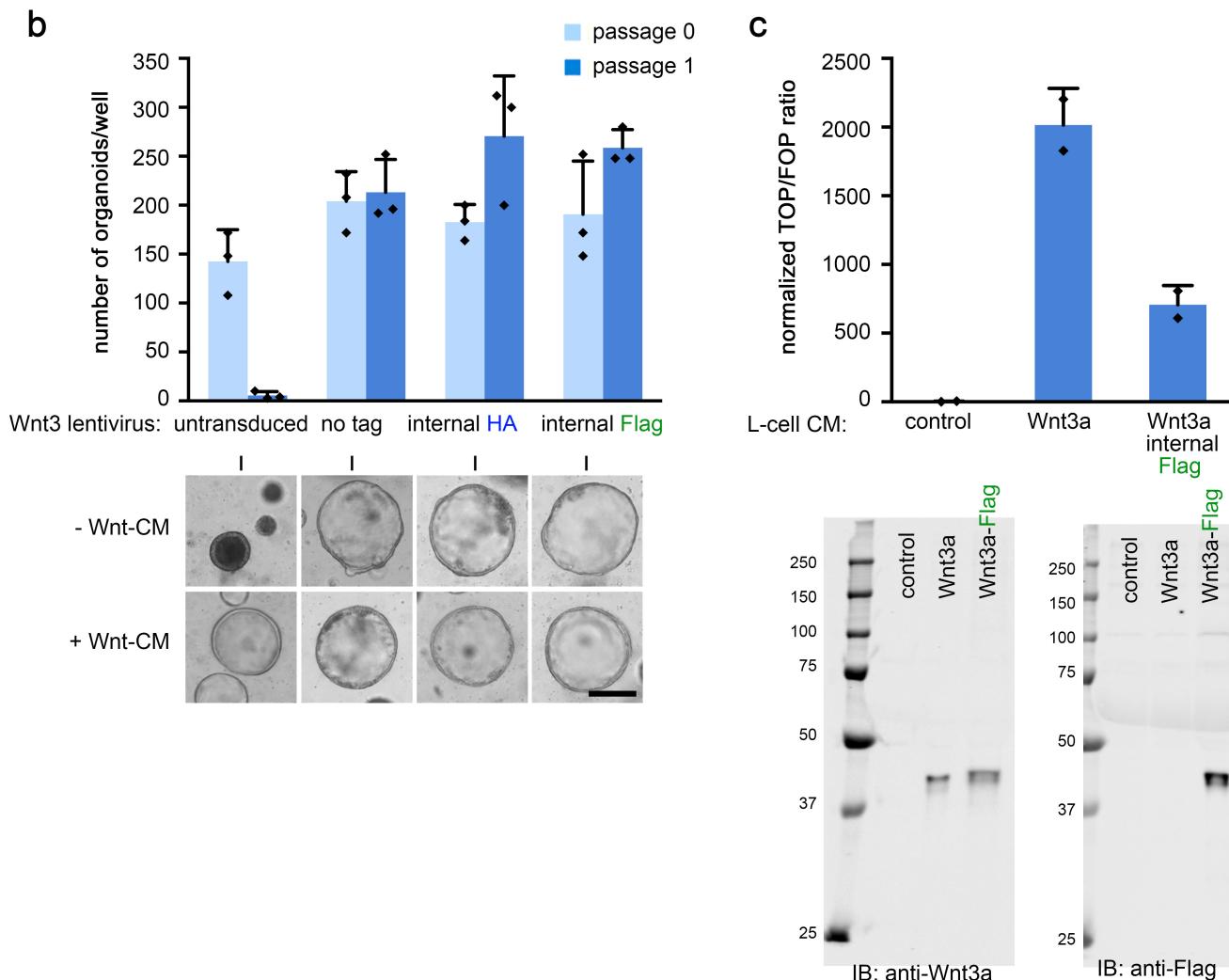


Extended Data Figure 1 | Generation of an epitope-tagged *Wnt3* mouse allele. **a**, Reported three-dimensional protein structure of Xwnt8 (in turquoise) bound to Fz8-CRD (in blue) (Protein Data Bank accession number 4FOA)¹⁹. Dotted line depicts the location of the amino (N) terminus, where the HA epitope tag was introduced. Note that the position of the tag does not interfere with receptor binding. The image was created with the jmol-viewer³⁷. Glycosyl and palmitoyl groups are shown in green and yellow, respectively. **b**, Amino-acid alignment of the N-terminal region of Wnt3 and other Wnt-family members. Predicted signal peptides are indicated in red. The HA tag (YPYDVPDYASL) was inserted after position Q41 that is labelled in blue. The first residue of Xwnt8 that is resolved in the crystal structure in **a** (T32) is shaded in green. Small letters indicate non-conserved residues that could not be aligned by the software³⁸. The degree of local similarity is marked by asterisks. A score of five asterisks represents maximal similarity, which is not reached using the default settings of the algorithm. **c**, HA-knock-in strategy: a targeting

vector that comprised the normal exon 2, a Frt-site flanked Pgk-neomycin resistance-polyA cassette and an inverted HA-inserted exon 2 (placed in intron 1 of the gene) was used. LoxP and LoxM sites were introduced in a configuration that allowed excision of the wild-type exon 2 and inversion of the HA-tagged exon 2 by Cre-mediated recombination (orientation of LoxP and LoxM sites is indicated with arrowheads). Transgenic mice were crossed to *Pgk-Cre* mice for germline recombination of the allele. This was done because the non-recombined allele was apparently a null allele as no homozygous offspring could be obtained. We suspect that the antisense configuration of the HA-modified exon 2 could result in RNA duplex formation and masking of the regular splice acceptor site of exon 2. Consistently, we observed a shorter transcript lacking wild-type exon 2 from the non-recombined allele by RT-PCR analysis. **d**, Southern blot analysis of ES cell clones to confirm correct targeting. Genomic BamHI digest was performed (see scheme in **c**).

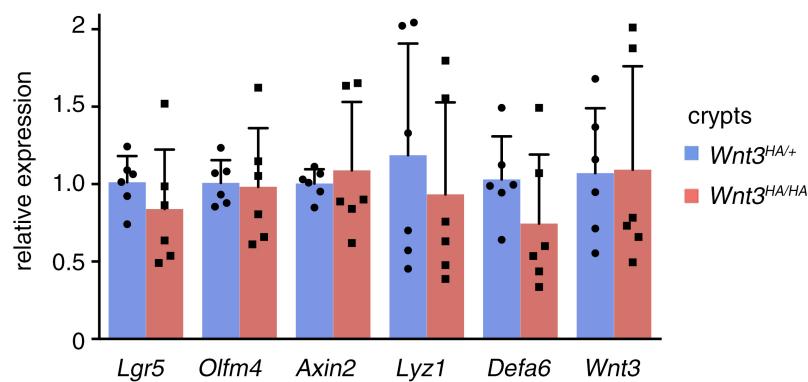
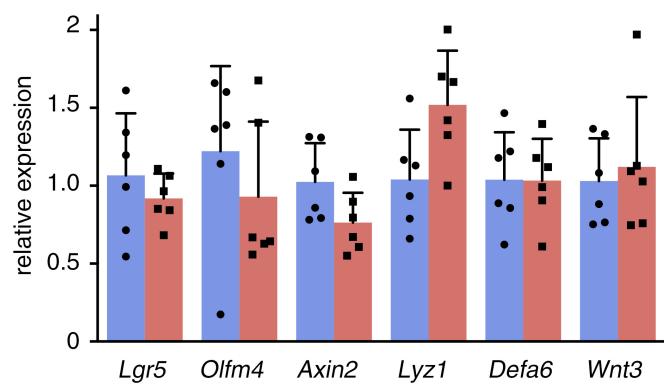
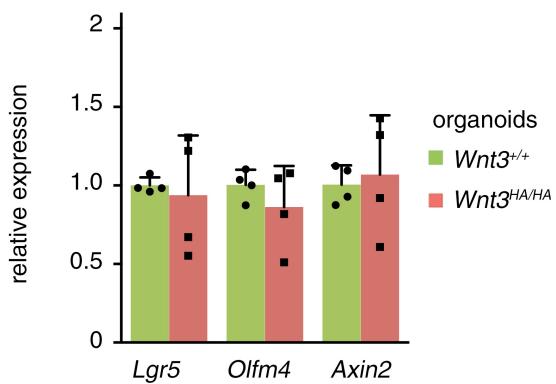
a

mWNT3	1	MEPHLLG LLL GG LL GG LL GG LL SGTRVLA G YPI W WS L ALGQQY T SLASQ-----PLLCGSI P GLVPKQLR FCR	60
mWNT3- HA	1	MEPHLLG LLL GG LL GG LL GG LL SGTRVLA G YPI W WS L ALGQQY T SLASQ Y PYDVPDY A SLPLLCGSI P GLVPKQLR FCR	71
mWNT3- Flag	1	MEPHLLG LLL GG LL GG LL GG LL SGTRVLA G YPI W WS L ALGQQY T SLASQ D YKDDDDK---PLLCGSI P GLVPKQLR FCR	68
mWNT3A	1	MAP--LG Y LL-VLCSLKQALG S YPI W SLAVGPQYSSL T Q-----PILCASIP G LVPKQLR FCR	57
mWNT3A- Flag	1	MAP--LG Y LL-VLCSLKQALG S YPI W SLAVGPQYSSL T Q D YKDDDDK---PILCASIP G GLVPKQLR FCR	65



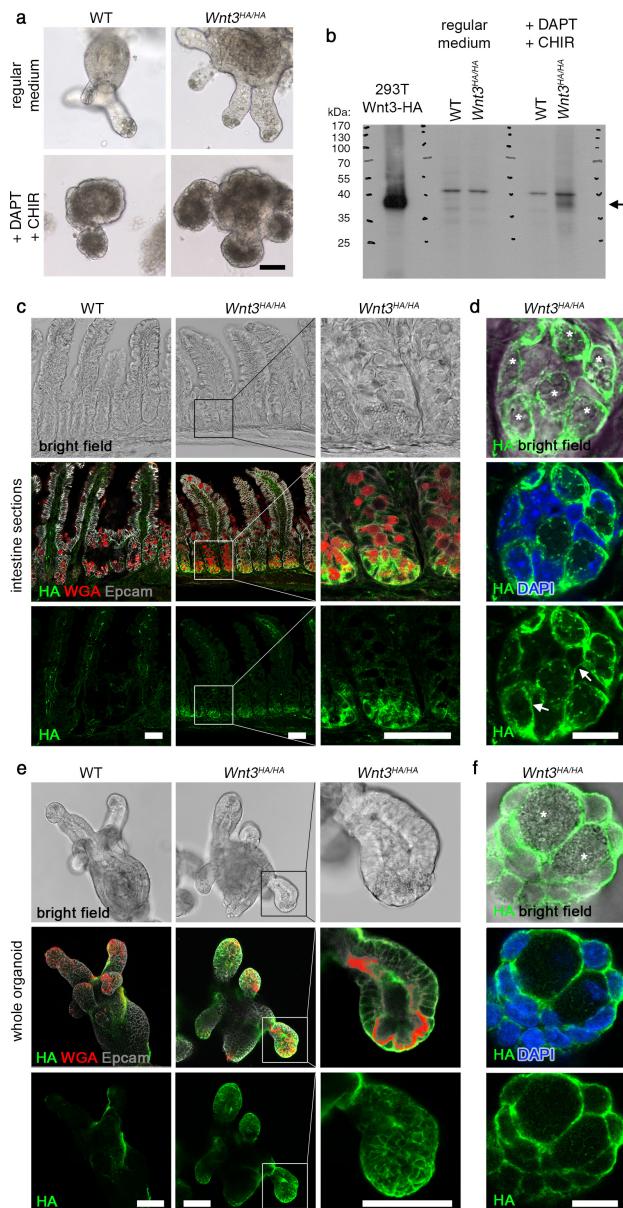
Extended Data Figure 2 | A permissive internal location for introduction of epitope tags in Wnt proteins. **a**, Amino-acid sequences of tagged Wnt versions used in this study. Signal peptides are labelled in red. Protein alignment (Clustal Omega program, Uniprot); asterisk, colon and dot symbols indicate full conservation, and groups of strongly and weakly similar properties, respectively. **b**, Lentiviral rescue experiment in Wnt-dependent ($Wnt3^{\Delta/\Delta}$) small intestinal organoids. Top: average organoid number (\pm s.d. in $n = 3$ independent wells) in the absence of exogenous Wnt-conditioned medium (Wnt-CM). Introduction of an

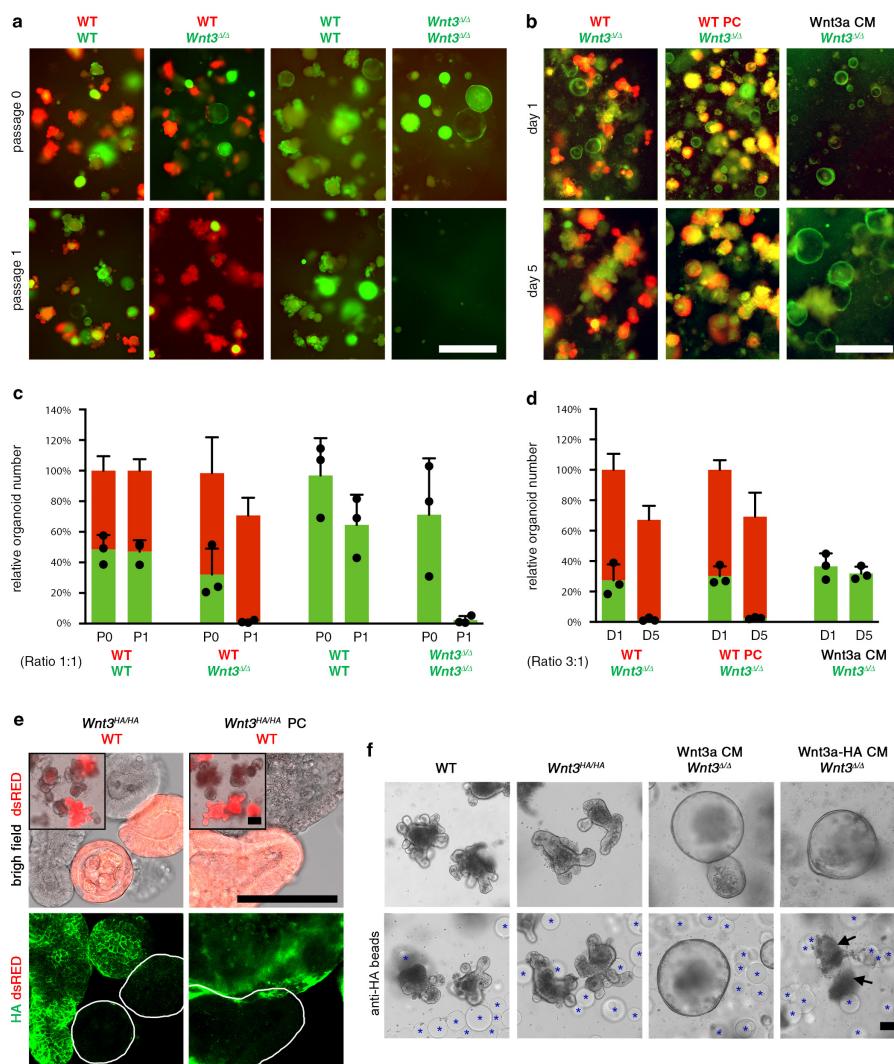
internal HA- or Flag-tag does not interfere with rescue activity of Wnt3 lentivirus. Bottom: cell morphology after 5 days (passage 0) in the absence or presence of Wnt-CM. Scale bar, 50 μ m. c, Introduction of an internal Flag-tag in mouse Wnt3a results in functional protein secretion in L-cells. Top: TOPFlash assay using conditioned media from control L-cells and stable lines expressing Wnt3a and Wnt3a-Flag. Bottom: immunoblots of the conditioned media measured above using anti-Wnt3a and anti-Flag antibodies. Fully scanned western blots are shown.

a**b****c**

Extended Data Figure 3 | Wnt signalling status in Wnt3-HA knock-in crypts and organoids. Quantitative RT-PCR analysis of freshly isolated crypts (a) or established organoids (b, c) from the small intestine. a, b, Mean normalized expression levels in Wnt3^{HA/HA} relative to Wnt3^{HA/+}-samples ($n=6$ mice, per genotype). c, Relative expression

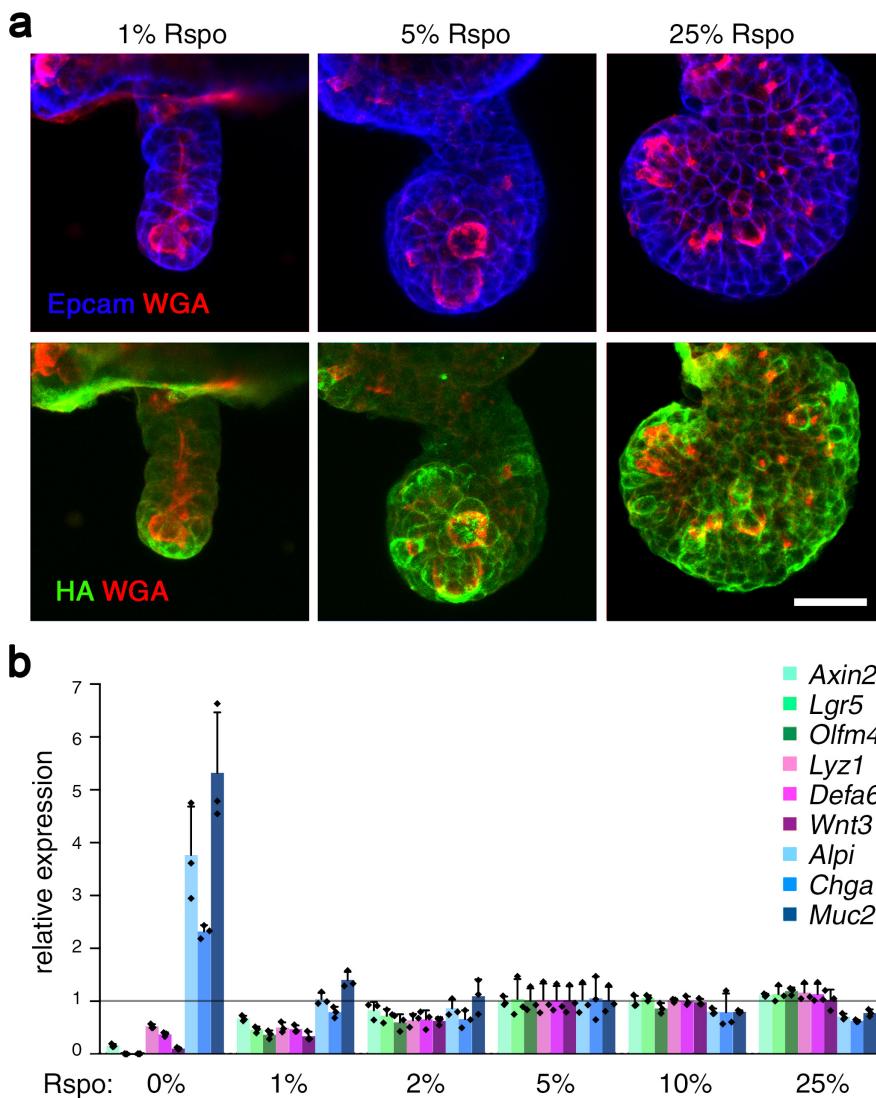
to Wnt3^{+/+} organoids ($n=4$ independent wells). Error bars, s.d., no significant changes were found ($P>0.05$ as determined by Student's *t*-test). Normal expression of stem-cell markers (*Lgr5*, *Olfm4*), Wnt pathway activity (*Axin2*) and Paneth cell markers (*Lyz1*, *Defa6* and *Wnt3*) indicates that introduction of the HA-tag does not interfere with Wnt signalling.





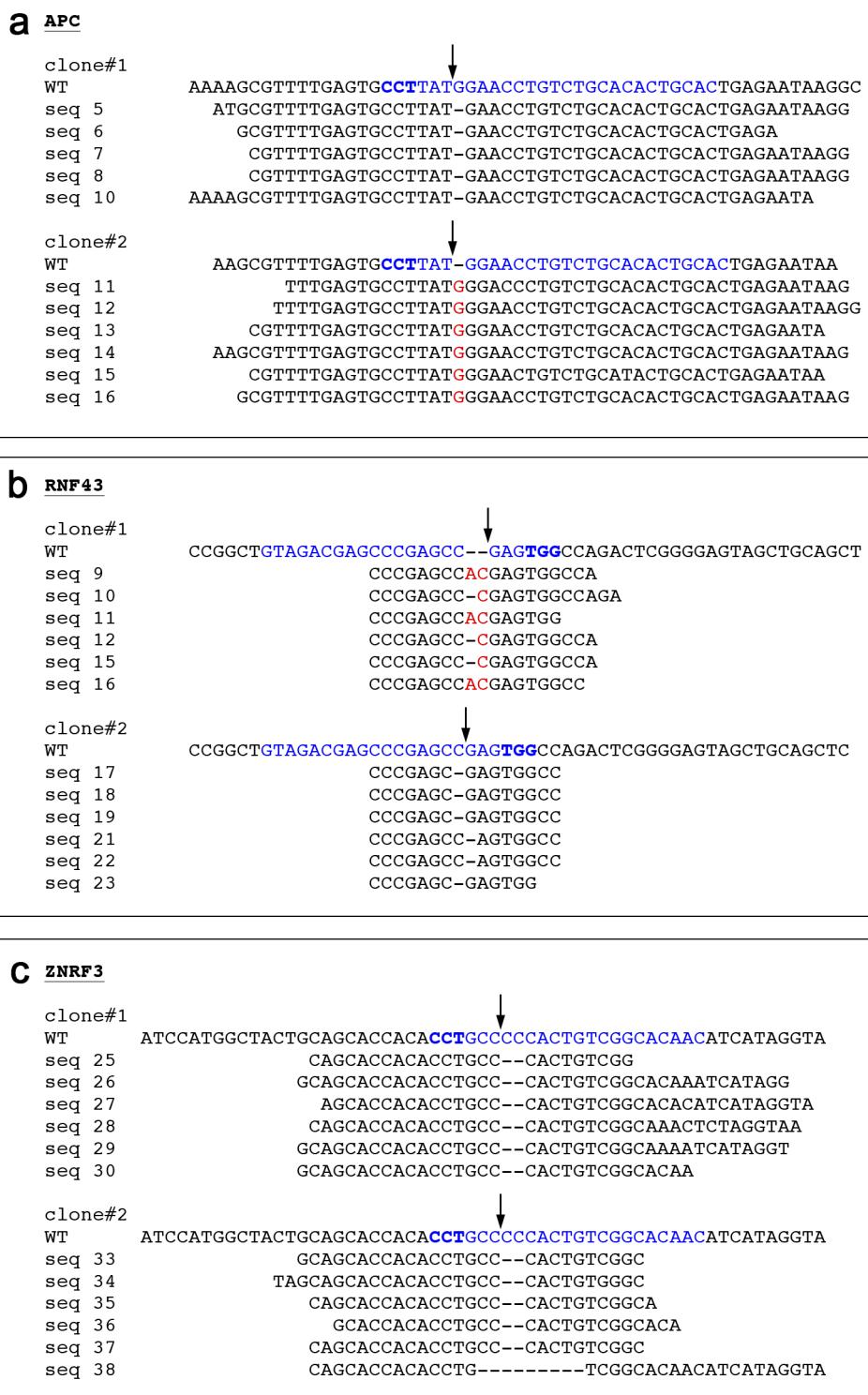
Extended Data Figure 5 | Diffusible Wnt activity in organoids is neither sufficient nor necessary to support growth. **a**, Culture after co-embedding of dsRED-labelled wild-type organoids with GFP-labelled wild-type or *Wnt3*^{Δ/Δ} organoids in Matrigel. Organoid fragments were seeded in a 1:1 ratio. *Wnt3*^{Δ/Δ} cells cannot be propagated alone or in the presence of wild-type cells. Images after seeding (P0; day 3) and after passage 1 (P1; day 10). **b**, Co-culture at higher seeding density using dsRED-labelled wild-type organoids or *in vitro* differentiated Paneth cells²¹ with GFP-labelled *Wnt3*^{Δ/Δ} organoids in a 3:1 ratio. Images of the same wells are shown at 1 and 5 days after seeding. **c, d**, Quantification of results shown in **a** and **b**. Mean relative number of organoids (\pm s.d.) from $n=3$ independent experiments. **e**, Anti-HA immunodetection following

co-culture of *Wnt3*^{HA/HA} organoids or *in vitro* differentiated Paneth cells with dsRED-labelled wild-type recipient organoids (3:1 ratio). Inserts show growth in Matrigel. Confocal z-projected images are shown. Note that dsRED-positive cells remain negative for Wnt3-HA. **f**, Bead depletion experiment. Wild-type, *Wnt3*^{HA/HA} or *Wnt3*^{Δ/Δ} organoids were either embedded Matrigel alone or together with anti-HA affinity beads (blue asterisks) to sequester diffusible HA-tagged Wnt. Bright-field images after 6 days of culture; note that *Wnt3*^{HA/HA} organoids display unaffected morphology in the presence of beads. *Wnt3*^{Δ/Δ} organoids do not grow in the presence of beads and L-cell derived Wnt3a-HA CM (black arrows) demonstrating efficient depletion. Scale bars, 500 μ m (**a, b**) and 100 μ m (**e, f**).



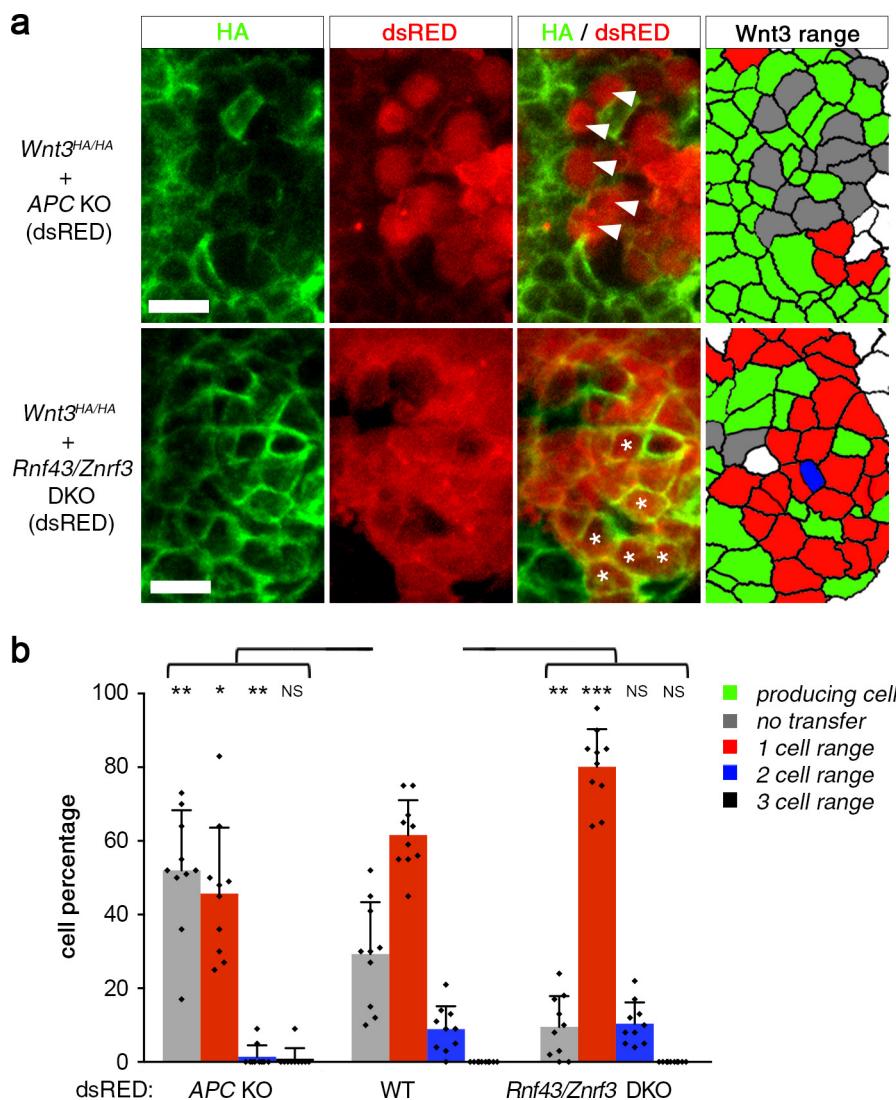
Extended Data Figure 6 | Wnt3 protein localization and messenger RNA expression depend on the R-spondin (Rspo) concentration.
a, Wnt3-HA immunostaining (*Wnt3*^{HA/HA} organoids) after culture for 6 days in 1%, 5% or 25% Rspo-conditioned medium. Counterstaining of plasma membranes (Epcam) and secretory granules (WGA). Scale bar, 25 μ m. **b**, RT-PCR analysis of stem-cell and differentiation markers in small intestinal organoids following 6 days of culture in variable

concentrations of Rspo. Shown are mean normalized expression levels (\pm s.d.) in $n = 3$ independent wells relative to organoids cultured in the normal concentration of 5% Rspo. Markers of stem cells (*Lgr5*, *Olfm4*), Wnt activity (*Axin2*), Paneth cells (*Lyz1*, *Defa6*, *Wnt3*), enterocytes (*Alpi*), endocrine cells (*Chga*) and goblet cells (*Muc2*). Note that the messenger RNA expression of *Wnt3* is sensitive to reduced Rspo concentration.



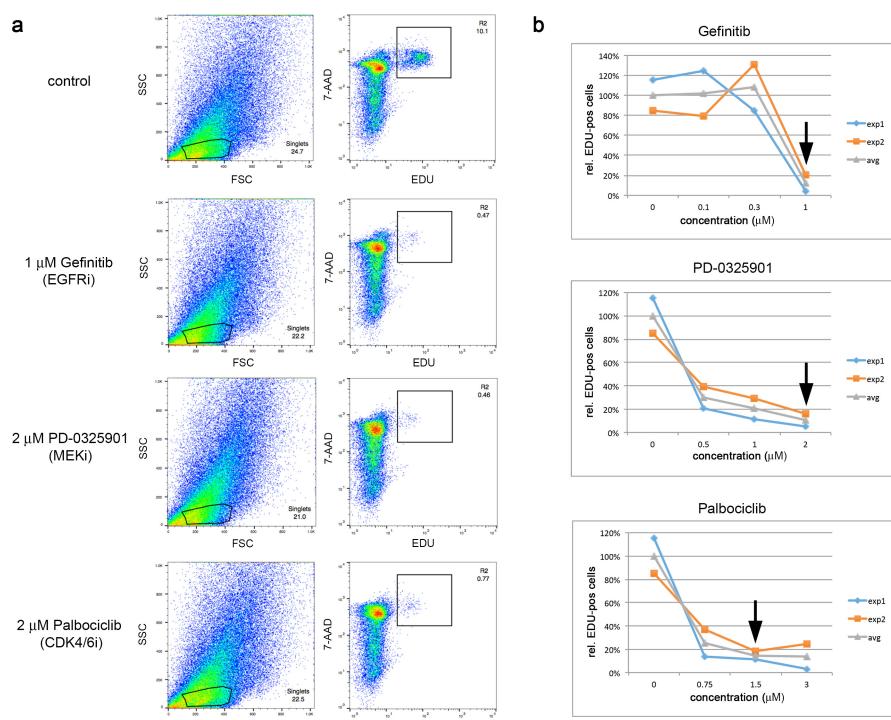
Extended Data Figure 7 | CRISPR/Cas9 induced mutations in organoids. a–c, Sequence analysis of indel mutations in targeted regions of mouse *APC* (a), *Rnf43* (b) and *Znrf3* (c) genomic loci. Two independent clonal organoid lines were analysed for each genotype. Genomic loci were PCR amplified, fragments were subcloned in *E. coli* and five or six

colonies were sequenced per clonal line. For *APC* (clone#1 and #2) and *Znrf3* (clone#1) mutant lines, hemizygosity was found, suggesting a larger genomic deletion on the other chromosome. sgRNA target sequences are shown in blue with the PAM sequence marked in bold. Arrows, Cas9 cleavage sites. Inserted nucleotides are shown in red.



Extended Data Figure 8 | Differential efficiency of Wnt3 transfer to *APC*- and *Rnf43/Znrf3*-deficient cells. **a**, Re-associated epithelia to test Wnt3 decoration on receiving cells (dsRED positive). The *z*-projected confocal images show depleted HA-signal on *APC* mutant cells (arrowheads) and enriched staining on *Rnf43/Znrf3* mutant cells (asterisks). Wnt3 range: receiving cells were colour-coded depending on

the success of transfer and the distance to the closest Wnt3-HA neighbour. **b**, Average percentages of each distance fraction in $n = 10$ re-associated organoids (\pm s.d.). Data were compared by *t*-test to re-association experiment using wild-type receiving cells (shown in Fig. 2f). $^{***}P < 10^{-3}$; $^{**}P < 10^{-2}$; NS, non-significant. Scale bars, $10\ \mu\text{m}$.



Extended Data Figure 9 | Pharmacological inhibition of proliferation in organoids. Cell cycle status of wild-type mouse small intestinal organoids was determined by flow cytometry. EDU incorporation assay in controls and 24 h after administration of EGFR-inhibitor (Gefinitib), MEK-inhibitor (PD-0325901) or CDK4/6-inhibitor (Palbociclib) to the regular culture medium. EDU was added 1 h before collection and dissociation of cells for FACS analysis. **a**, Original FACS data. Single cells were gated using

FSC/SSC characteristics. EDU signals were plotted against 7-AAD signals (DNA content). **b**, Inhibitor titration. Relative percentage of EDU-positive cells is shown for two independent experiments (grey line shows average). Experiments in Fig. 4 were performed at concentrations indicated by arrows. Note that Lgr5⁺ intestinal stem cells double every 19–20 h *in vitro*²¹, comparable to the cell cycle length *in vivo*³⁹ (21.5 h).

The sexual identity of adult intestinal stem cells controls organ size and plasticity

Bruno Hudry¹, Sanjay Khadayate¹ & Irene Miguel-Aliaga¹

Sex differences in physiology and disease susceptibility are commonly attributed to developmental and/or hormonal factors, but there is increasing realization that cell-intrinsic mechanisms play important and persistent roles^{1,2}. Here we use the *Drosophila melanogaster* intestine to investigate the nature and importance of cellular sex in an adult somatic organ *in vivo*. We find that the adult intestinal epithelium is a cellular mosaic of different sex differentiation pathways, and displays extensive sex differences in expression of genes with roles in growth and metabolism. Cell-specific reversals of the sexual identity of adult intestinal stem cells uncovers the key role this identity has in controlling organ size, reproductive plasticity and response to genetically induced tumours. Unlike previous examples of sexually dimorphic somatic stem cell activity, the sex differences in intestinal stem cell behaviour arise from intrinsic mechanisms that control cell cycle duration and involve a new *doublesex*- and *fruitless*-independent branch of the sex differentiation pathway downstream of *transformer*. Together, our findings indicate that the plasticity of an adult somatic organ is reversibly controlled by its sexual identity, imparted by a new mechanism that may be active in more tissues than previously recognized.

Sex differences in intestinal physiology³ prompted us to investigate possible molecular underpinnings. RNA-seq transcriptional profiling of virgin adult midguts (GEO accession number GSE74775) indicated significant sexual dimorphism in gene expression and/or splicing, with 8.3% of all genes and 5.6% of all isoforms expressed in the midgut displaying differences in expression between the sexes (Extended Data Fig. 1a–f and Supplementary Table 1). Sex-biased expression or splicing was confirmed for a subset of genes by real time qRT-PCR (Extended Data Fig. 1h, i). Genes with sex differences in expression cluster into distinct biological processes; genes assigned to cell division-related processes are more abundantly expressed in females, whereas genes coding for proteins that function in carbohydrate metabolism and redox processes are preferentially expressed in males (Extended Data Fig. 1g).

The above results suggested active expression of sex determinants in the adult intestine. In *Drosophila*, sex chromosome sensing leads to sexually dimorphic expression of *Sex lethal* (*Sxl*): the master regulator of both sexual development and dosage compensation (DC, an epigenetic process by which transcription of the single male X chromosome is upregulated twofold⁴). *Sxl* controls the sex-specific splicing of its downstream target gene *transformer* (*tra*)⁵, leading to functional *Tra* protein expression only in females, and to the sex-specific splicing of two *Tra* direct targets—the transcription factors *doublesex* (*dsx*) and *fruitless* (*fru*)—that sculpt sexually dimorphic anatomical features, reproductive systems and behaviour^{6–8} (Extended Data Fig. 2b). As may be expected from their roles in DC⁴, *Sxl* and the sexually dimorphic DC complex (DCC) member Male-specific lethal 2 (Msl-2) are ubiquitously expressed in the intestinal epithelium of females and males, respectively, in both adult intestinal progenitors (intestinal stem cells (ISCs) and postmitotic enteroblasts (EBs)) and their differentiated progeny

(enterocytes (ECs) and enteroendocrine cells (EECs)) (Extended Data Fig. 2a–d). Consistent with our RNA-seq analysis, a newly generated *tra-LacZ* reporter is also expressed in all four epithelial cell types of the adult intestine (Extended Data Fig. 2a, e), suggesting that downstream of *Sxl*, the sex differentiation pathway remains active in this adult epithelium. However, the two targets of *Tra*-mediated sex-specific splicing are either absent from all midgut epithelial cells (*Fru*^M) or are expressed in only a subset of epithelial cells (*Dsx*^{F/M}, expressed in ECs but absent from ISCs, EBs and EECs) (Extended Data Fig. 2a, f, g). Hence, the expression of sex determination genes is maintained in the adult midgut, but displays cell-type specificity; while adult-born enterocytes express all members of the canonical sex determination pathway, their siblings (the EECs) and both types of adult intestinal progenitors (ISCs and EBs) express the early (*Sxl*, *tra*), but not the late (*Dsx*^{F/M}, *Fru*^M), effectors of the pathway.

Both the enrichment analysis and the presence of *Sxl/tra* in adult ISCs pointed to sexually dimorphic ISC proliferation. Female flies exhibit a rapid proliferative response to dextran sodium sulphate (DSS)-induced damage of the intestinal epithelium (Fig. 1a and ref. 9). This response was less pronounced in male midguts (Fig. 1a) or in female (but not male) midguts following adult-restricted *Sxl* downregulation in intestinal progenitors (Fig. 1a and Extended Data Fig. 3a, c). Conversely, ectopic expression of *Sxl* in adult intestinal progenitors enhanced proliferation in male (but not female) midguts (Fig. 1a and Extended Data Fig. 3a). Additional cell-type- and adult-specific *Sxl* downregulation experiments indicated that *Sxl* acts in ISCs, and not in other cells, to control sexually dimorphic proliferation rather than differentiation (Fig. 1e and Extended Data Figs 3d and 4a). Mechanistically, females do not have a significantly higher density of ISCs than male flies (Fig. 1b) or a higher ratio of symmetric versus asymmetric divisions (Fig. 1c), suggesting that the proliferative capacity of female ISCs is intrinsically enhanced by their expression of *Sxl*. Consistent with this idea, a higher percentage of their adult progenitors are found in G2/S phase at the expense of G1 in homeostatic conditions (Fig. 1d), suggestive of shorter cell cycles, and adult-specific downregulation of *Sxl* in intestinal progenitors abrogated the sexual dimorphism in G2/S to G1 ratio, without affecting the number of ISCs or their division mode (Figs 1b–d). Clonal analyses further confirmed the intrinsic nature of the sexual dimorphism in proliferation, its *Sxl* control and adult reversibility both during regeneration and homeostasis (Fig. 1g and Extended Data Fig. 3b, e and f).

To investigate whether the reported *Sxl* effects result from deregulated DC, we first confirmed that DC can be functionally inactivated in adults by observing loss of histone H4 lysine 16 acetylation of the X chromosome upon adult-specific downregulation of *msl-2* in male intestinal progenitors (Extended Data Fig. 3g). We then investigated whether ectopic *msl-2* expression accounted for the reduced proliferation resulting from *Sxl* downregulation by co-downregulating both genes in adult intestinal progenitors. This did not reinstate female proliferation (Fig. 1f). The converse experiment—misexpression of

¹MRC Clinical Sciences Centre, Imperial College London, Hammersmith Campus, Du Cane Road, London W12 0NN, UK.

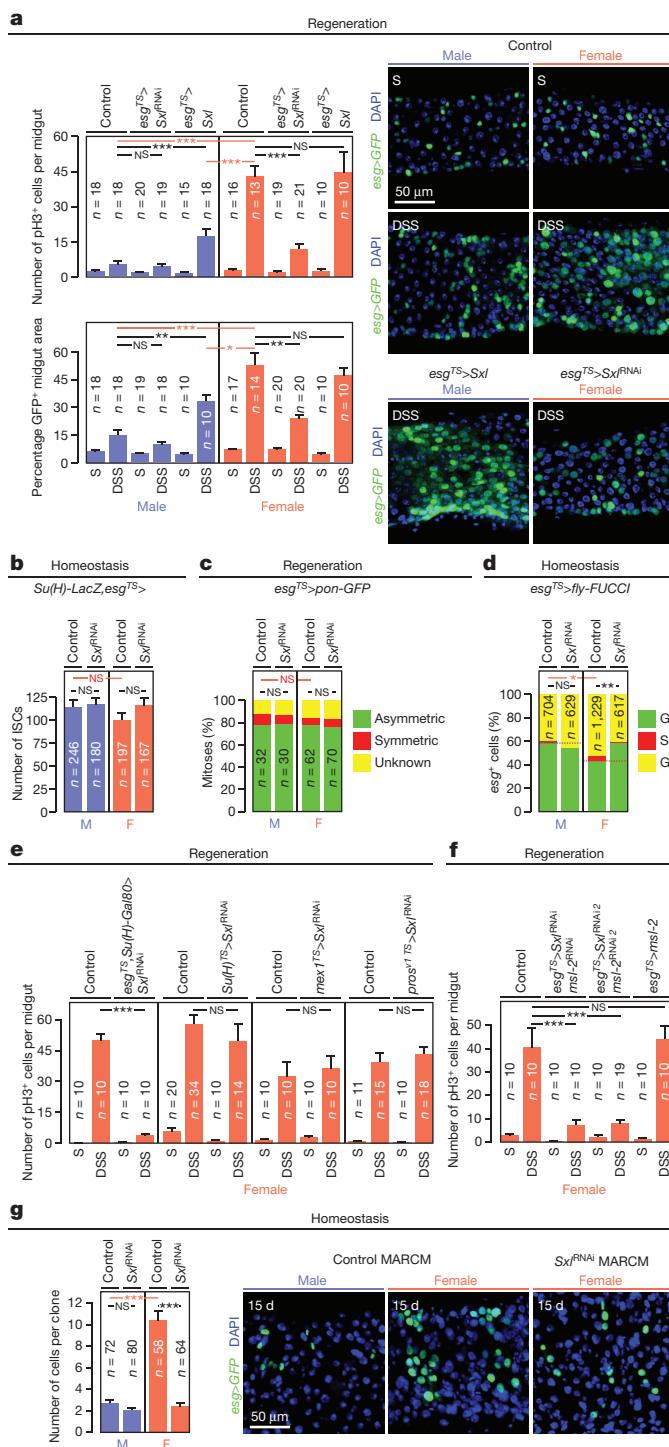


Figure 1 | *Sxl* controls intrinsic sex differences in adult ISC proliferation independently of dosage compensation. **a**, Number of mitoses (phospho-histone H3 (pH3)-positive cells, top graph) and percentage of cells positive for an intestinal progenitor marker (*escargot* (*esg*)-positive/total area, bottom graph) in controls and flies with adult-restricted downregulation or misexpression of *Sxl* in ISCs/EBs (achieved by *esg-Gal4*, *tub-Gal80^{TS}*-driven *Sxl* RNAi or *UAS-Sxl*, respectively). Flies were exposed to control (sucrose, S) or damage-inducing (DSS) diets. Representative images are shown to the right (DNA: DAPI, in blue; ISC/EB marker: GFP, in green). **b**, Stem cell number (*esg⁺*, *Suppressor of Hairless* (*Su(H)*)⁻ cells) in the posterior midgut following 20 days of adult ISC/EB-specific *Sxl* downregulation. **c**, Quantifications of symmetric (red) versus asymmetric (green) ISC divisions based on cortical *Partner of Numb* (*Pon*)-GFP distribution in metaphase and telophase reveal no differences between the sexes or upon adult ISC/EB-specific *Sxl* downregulation. Mitoses with lack of clear *Pon*-GFP signal are displayed in yellow. **d**, Percentage of progenitors in G1, S or G2 as revealed by ISC/EB-driven expression of the cell cycle indicator Fly-FUCCI. **e**, Number of mitoses in DSS-treated flies with adult-specific downregulation of *Sxl* in ISCs (*esg-Gal4*, *Su(H)-Gal80* driver), EBs (*Su(H)-Gal4*), ECs (*midgut expression 1 (mex1)-Gal4*) or EECs (*prospero (pros)^{V1}-Gal4*). **f**, pH3 quantifications following adult-specific and simultaneous *msl-2/Sxl* downregulation or *msl-2* misexpression in ISCs/EBs. **g**, MARCM clone size quantifications (graph) and representative images (labelled in green with GFP) reveal that clones expressing *Sxl*^{RNAi} are smaller than control clones in females, but not in males. *n* denotes the number of midguts (**a**, **e**, **f**), ISCs/EBs (**b**, **d**), mitoses (**c**) or clones (**g**) that were analysed for each genotype. Results were combined from at least two independent experiments. See Supplementary Information for full genotypes. In this and all subsequent figures, error bars correspond to standard error of the mean (s.e.m.).

requirement for *tra* in regulating sexually dimorphic proliferation, both during regeneration (Extended Data Fig. 5b, d) and in normal homeostasis (Fig. 2c and Extended Data Fig. 5e, f). Strikingly, reintroduction of a *tra* transgene specifically in adult intestinal progenitors fully rescued the reduced proliferation resulting from *Sxl* downregulation (Fig. 2b). Together with the *Sxl* experiments, these results show that the sex of the midgut is actively specified in adult flies. Unlike other adult somatic cell types^{6,10,11}, adult ISCs have a plastic sexual identity manifested by an intrinsic, *Sxl/tra*-controlled and DC-independent sexual dimorphism in both their basal and regenerative proliferation. These findings extend previous work in both mouse and *Drosophila*^{12–16} by showing that sexual identity not only needs to be actively maintained, but can also be reversed bidirectionally in the adult cells of a non-gonadal organ.

Unexpectedly (and in contrast to *tra*), mutation of the canonical Tra binding partner Transformer 2 (Tra2)^{5,7} failed to affect the sex differences in ISC proliferation (Fig. 2d), despite interfering with the sex-specific expression of *dsx^F* and *Yolk protein 1 (Yp1)* transcripts as anticipated (Extended Data Fig. 5g). Together with our finding that ISCs do not express *Dsx* or *Fru^M* (Extended Data Fig. 2a, f, g), this result suggested that a non-canonical Tra2-, *Dsx*-, *Fru^M*-independent sex determination pathway drives sexually dimorphic proliferation in adult intestinal stem cells. We conducted a series of rescue, gain- and loss-of-function experiments using *dsx^F* transgenes, *dsx/fru* mutants, RNAi transgenes and mutants in which *dsx/fru* splicing was shifted towards either male or female isoforms (Extended Data Fig. 6a–g). These ruled out both a direct action of *dsx* or *fru^M* in intestinal progenitors as well as indirect contributions to ISC proliferation from other *dsx/fru^M*-expressing cells such as the neighbouring ECs, thus confirming that the sex of adult somatic stem cells in the intestine is specified by a novel branch of the sex determination pathway.

We then combined new RNA-seq and genetic experiments to identify *tra* target genes in adult ISCs, defined as genes of which the expression or splicing was different between control and *tra* mutant females, but returned to levels comparable to those of controls upon adult ISC-specific re-introduction of *tra* (see Methods, GEO accession number GSE74775 and Supplementary Table 1 for details). This yielded 72 genes with *tra*-regulated expression (34) or splicing (38) in

msl-2 in adult female intestinal progenitors using a newly generated transgene coding for HA-tagged Msl-2—did not reduce their proliferation (Fig. 1f) despite efficient Msl-2 protein expression and function (Extended Data Fig. 3h and data not shown). Hence, DC does not account for the ISC sex differences.

This focused our attention on the sex differentiation pathway and its main effector *tra*. Like *Sxl*, *tra* downregulation reduced DSS-induced proliferation in females to levels comparable to those seen in male midguts, but did not affect proliferation in male midguts (Fig. 2a and Extended Data Fig. 5a). Conversely, *tra* misexpression, either ubiquitous (Extended Data Fig. 5c) or confined to adult intestinal progenitors (Fig. 2a and Extended Data Fig. 5a), increased the proliferative response of ISCs to DSS in adult males, but not in females. Clonal and *tra* mutant rescue experiments further confirmed the adult, cell-intrinsic

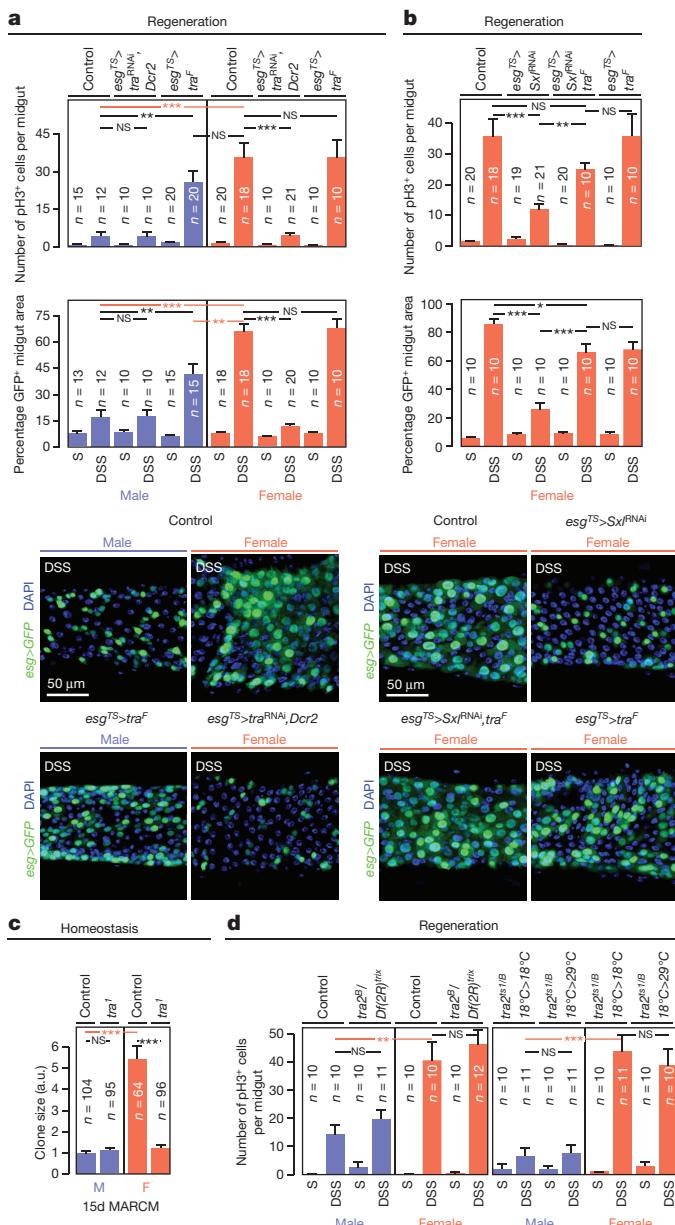


Figure 2 | *tra*, but not *tra2*, controls intrinsic sex differences in adult ISC proliferation. **a**, Graphs of mitoses (pH3+, top) and intestinal progenitor area (*esg*>GFP, bottom) in flies exposed to control (sucrose, S) or damage-inducing (DSS) diets, in both controls and flies with adult ISC/EB-restricted *tra* downregulation or misexpression. Representative images are shown below the graphs (DNA: DAPI, in blue; ISC/EB marker: GFP, in green). **b**, Comparable quantifications of the *Sxl* downregulation phenotypes in females, and their rescue by re-expression of *tra*^F. **c**, Clone size quantifications (in arbitrary units (a.u.) of GFP fluorescence, see Methods) reveal that *tra* null mutant MARCM clones are smaller than control clones only in females. **d**, No significant differences in ISC proliferation in the midguts of DSS-treated males or females lacking *tra2* (*tra2*^{B/D(f2R)trix}), or lacking *tra2* specifically in adults (achieved by shifting flies with the thermosensitive allele *tra2*^{ts1} from 18 °C to 29 °C in the adult stage) versus controls. *n* denotes the number of midguts (a, e, f), ISC/EBs (b, d), mitoses (c) or clones (g) that were analysed for each genotype. Results were combined from at least two independent experiments. See Supplementary Information for full genotypes.

adult intestinal progenitors (Fig. 3a, b and Extended Data Fig. 7a–c). A genetic screen (Extended Data Fig. 7d, see Methods for details) identified three of them as negative or positive regulators of proliferation accounting for the sexual dimorphism in ISC proliferation. Indeed, downregulation of *Imaginal disc growth factor 1* (*Idgf1*) or *Serpin 88E*

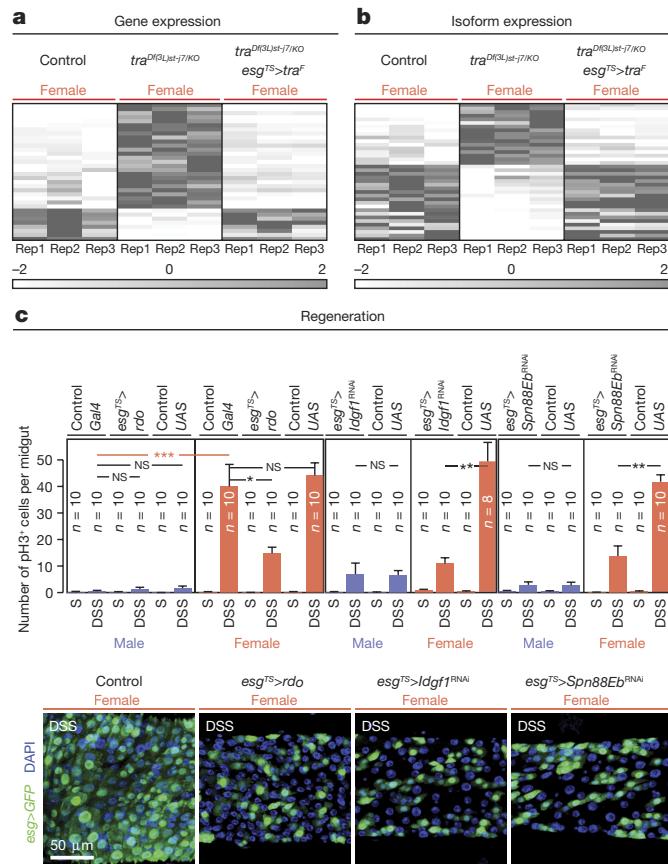


Figure 3 | *tra* targets in adult intestinal progenitors. **a**, **b**, Heat maps of the genes with *tra*-regulated expression (a) or splicing (b) in ISC/EBs, displaying their normalized abundance in females, *tra* null mutant females and *tra* null mutant females with feminized ISC/EBs (*esg*^{TS}>*tra*^F) (see Methods for details). **c**, Mitoses (pH3) quantifications following adult ISC/EB-confined manipulations of *rdo*, *Idgf1* and *Spn88Eb* expression. Representative images for each genotype are shown below the graphs (DNA: DAPI, in blue; ISC/EB marker: GFP, in green). *n* = 10 midguts per genotype/condition. Results were combined from at least two independent experiments. See Supplementary Table 1 for a full list of names and quality scores, and Supplementary Information for full genotypes.

(*Spn88Eb*), normally activated by *tra* in adult intestinal progenitors, reduced proliferation only in females. By contrast, overexpression of *reduced ocelli* (*rdo*), normally repressed by *tra* in the same cells, reduced proliferation only in females (Fig. 3c and Extended Data Fig. 7e). These three targets are regulated at the expression—rather than splicing—level, consistent with novel, *tra2*- and, possibly, splicing-independent roles of *tra*. Once produced, they may be secreted or targeted to the cell surface^{17–19}, suggesting that sex differences may result from modulation of how ISCs interact with their local environment. All three proteins also appear to be expressed in other tissues and/or during development, and belong to protein families also represented in mammals: leucine-rich repeat (*rdo*), inhibitory serpins (*Spn88Eb*) and secreted glycoproteins (*Idgf1*)^{17,19,20}. It will therefore be interesting to explore whether they account for how organisms attain their sexually dimorphic body size during development.

Are the sex differences in ISCs physiologically significant? Wild-type female midguts are longer than male midguts at days 3, 5 and 20 of adult life (Fig. 4a and data not shown). Adult-specific and cell-autonomous masculinization of ISCs (achieved by *tra* or *Sxl* downregulation) dramatically shrank the female (but not the male) midgut towards a size more similar to that of males (Fig. 4a and Extended Data Fig. 8c). Cell type- and region-specific quantifications indicated that a reduction in ISC proliferation and, consequently, the number of differentiated progeny accounted for the observed shrinkage (Extended Data

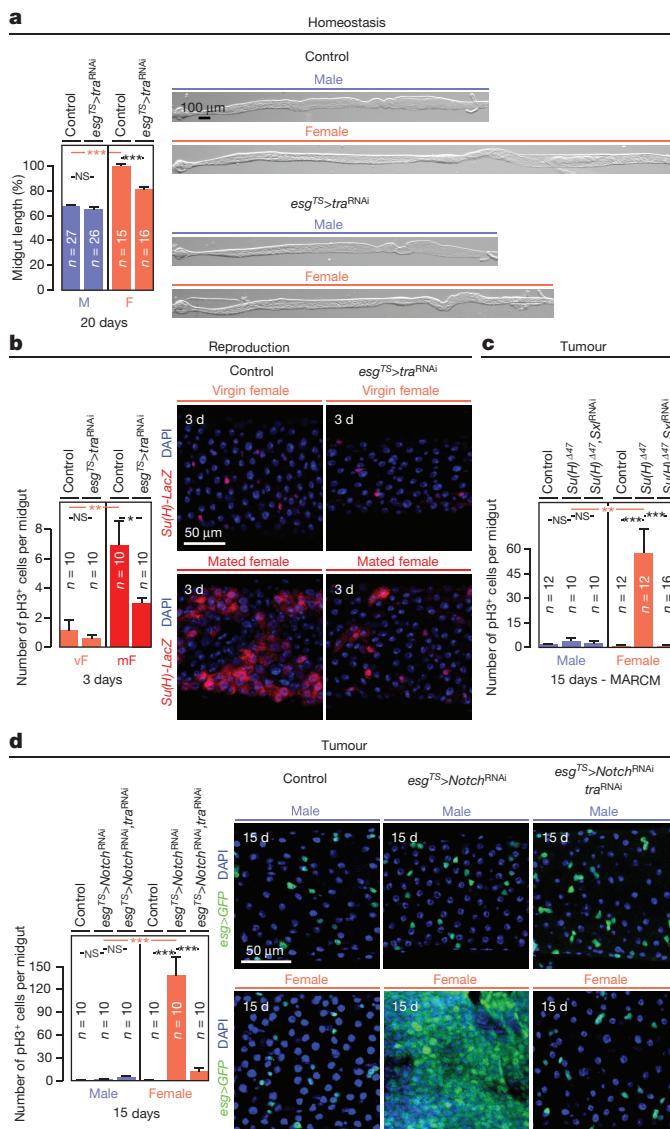


Figure 4 | Physiological importance of the sex differences in intestinal progenitors. **a**, Midgut length quantifications and representative images of phenotypes resulting from adult ISC/EB-specific masculinization of ISCs (achieved by *esg*^{TS}-driven *tra* downregulation initiated after the phase of midgut post-eclosion growth, see Methods for details). **b**, The number of mitoses (pH3-positive cells) is higher in control female flies 3 days after mating. The postmating increase is abrogated upon adult ISC/EB-specific *tra* downregulation. An EB marker (*Su(H)*-LacZ, in red in image panels) reveals that the EB expansion seen in females after mating is reduced upon adult ISC/EB-specific *tra* downregulation. See also Extended Data Fig. 8 for quantifications. **c**, pH3 quantifications inside MARCM clones of control flies, *Su(H)* mutants and *Su(H)* mutants in which *Sxl* has been downregulated inside the clone. *Su(H)* mutation only leads to increased pH3 counts in females, and this increase is *Sxl* dependent. **d**, Hyperplasia (quantified by the number of pH3-positive cells) resulting from adult ISC/EB-driven *Notch* downregulation and its modulation by *tra* in female and male midguts. Confocal images show intestinal progenitor coverage of representative midgut portions for each genotype (DNA: DAPI, in blue; ISC/EB marker: GFP, in green). *n* denotes the number of midguts (**a**, **b**, **c**, **d**) that were analysed for each genotype. Virgin flies were used in all experiments unless otherwise indicated. Results were combined from at least two independent experiments. See Supplementary Information for full genotypes.

Fig. 8a–d). Female flies with masculinized adult ISCs also failed to undergo the recently described midgut resizing triggered by mating²¹ (Fig. 4b and Extended Data Fig. 8e) and had slightly reduced fecundity (Extended Data Fig. 8f), indicating that sex differences endow females

with enhanced stem cell plasticity to optimise reproduction. Females are, however, more prone to genetically induced tumours; indeed, adult-specific interference with *Notch* (*N*) or *Apc-ras* signalling, previously shown to lead to tissue overgrowth reminiscent of gastrointestinal tumours in females^{22,23}, resulted in hyperplasia of female (but not male) midguts (Fig. 4c, d and Extended Data Fig. 9a–f). Hyperplasia resulted from sex differences in proliferation rather than sex-specific differentiation defects (Extended Data Fig. 4b), and could be prevented by simultaneously masculinizing female adult ISCs by downregulating or mutating *Sxl/tra*, but not *dsx* and *fru*^M (Fig. 4c, d and Extended Data Fig. 9a, c–g). Increased susceptibility of female flies to genetically induced tumours was also observed after mating (data not shown). Thus, the intrinsic sexual identity of adult intestinal stem cells plays key roles in adult life, both in maintaining organ size and in modulating its plasticity.

Previous observations had pointed to additional branches of the sex determination pathway^{24,25}. Our discovery of this new branch has important implications for organs such as the nervous system, where sexual identity was thought to be confined to *Fru*^M- and/or *Dsx*-expressing neurons²⁶, and raises the possibility that every cell has a sexual identity that actively regulates its plasticity and physiology. Although early sex specification in *Drosophila* differs from that of mammals, there is increasing evidence that it converges on common effectors such as the *Dsx/Dmrt* family of transcription factors and their targets¹². Hence, the recently reported sex differences in both intestinal gene expression and microbiota in the mammalian intestine^{27,28} could at least partly result from similar, so far unexplored, intrinsic sex differences. Similarly, the possible contribution of the intrinsic sexual identity of adult somatic stem cells to both organ plasticity and to the well documented sex differences in susceptibility to many types of cancer^{29,30} deserves further investigation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 29 June; accepted 21 December 2015.

- Arnold, A. P. The end of gonad-centric sex determination in mammals. *Trends Genet.* **28**, 55–61 (2012).
- Ober, C., Loisel, D. A. & Gilad, Y. Sex-specific genetic architecture of human disease. *Nature Rev. Genet.* **9**, 911–922 (2008).
- Cognigni, P., Bailey, A. P. & Miguel-Aliaga, I. Enteric neurons and systemic signals couple nutritional and reproductive status with intestinal homeostasis. *Cell Metab.* **13**, 92–104 (2011).
- Lucchesi, J. C. & Kuroda, M. I. Dosage compensation in *Drosophila*. *Cold Spring Harb. Perspect. Biol.* **7**, <http://dx.doi.org/10.1101/cshperspect.a019398> (2015).
- Boogs, R. T., Gregor, P., Idriss, S., Belote, J. M. & McKeown, M. Regulation of sexual differentiation in *D. melanogaster* via alternative splicing of RNA from the transformer gene. *Cell* **50**, 739–747 (1987).
- Camara, N., Whitworth, C. & Van Doren, M. The creation of sexual dimorphism in the *Drosophila* soma. *Curr. Top. Dev. Biol.* **83**, 65–107 (2008).
- Christiansen, A. E., Keisman, E. L., Ahmad, S. M. & Baker, B. S. Sex comes in from the cold: the integration of sex and pattern. *Trends Genet.* **18**, 510–516 (2002).
- Villella, A. & Hall, J. C. Neurogenetics of courtship and mating in *Drosophila*. *Adv. Genet.* **62**, 67–184 (2008).
- Amcheslavsky, A., Jiang, J. & Ip, Y. T. Tissue damage-induced intestinal stem cell division in *Drosophila*. *Cell Stem Cell* **4**, 49–61 (2009).
- Arthur, B. I. Jr, Jallon, J. M., Caflisch, B., Choffat, Y. & Nothiger, R. Sexual behaviour in *Drosophila* is irreversibly programmed during a critical period. *Current Biol.* **8**, 1187–1190 (1998).
- Ferveur, J. F. et al. Genetic feminization of pheromones and its behavioral consequences in *Drosophila* males. *Science* **276**, 1555–1558 (1997).
- Clough, E. et al. Sex- and tissue-specific functions of *Drosophila* doublesex transcription factor target genes. *Dev. Cell* **31**, 761–773 (2014).
- Ma, Q., Wawersik, M. & Matunis, E. L. The Jak–STAT target Chinmo prevents sex transformation of adult stem cells in the *Drosophila* testis niche. *Dev. Cell* **31**, 474–486 (2014).
- Matson, C. K. et al. DMRT1 prevents female reprogramming in the postnatal mammalian testis. *Nature* **476**, 101–104 (2011).
- Shapiro-Kulnane, L., Smolko, A. E. & Salz, H. K. Maintenance of *Drosophila* germline stem cell sexual identity in oogenesis and tumorigenesis. *Development* **142**, 1073–1082 (2015).
- Uhlenhaut, N. H. et al. Somatic sex reprogramming of adult ovaries to testes by FOXL2 ablation. *Cell* **139**, 1130–1142 (2009).

17. Caldwell, J. C., Fineberg, S. K. & Eberl, D. F. *reduced ocelli* encodes the leucine rich repeat protein *Pray For Elves* in *Drosophila melanogaster*. *Fly* **1**, 146–152 (2007).
18. Handke, B. *et al.* The hemolymph proteome of fed and starved *Drosophila* larvae. *PLoS ONE* **8**, e67208 (2013).
19. Kawamura, K., Shibata, T., Saget, O., Peel, D. & Bryant, P. J. A new family of growth factors produced by the fat body and active on *Drosophila* imaginal disc cells. *Development* **126**, 211–219 (1999).
20. Brown, J. B. *et al.* Diversity and dynamics of the *Drosophila* transcriptome. *Nature* **512**, 393–399 (2014).
21. Reiff, T. *et al.* Endocrine remodelling of the adult intestine sustains reproduction in *Drosophila*. *eLife* **4**, e06930 (2015).
22. Ohlstein, B. & Spradling, A. The adult *Drosophila* posterior midgut is maintained by pluripotent stem cells. *Nature* **439**, 470–474 (2006).
23. Patel, P. H. & Edgar, B. A. Tissue design: how *Drosophila* tumors remodel their neighborhood. *Semin. Cell Dev. Biol.* **28**, 86–95 (2014).
24. Evans, D. S. & Cline, T. W. *Drosophila* switch gene *Sex-lethal* can bypass its switch-gene target *transformer* to regulate aspects of female behavior. *Proc. Natl Acad. Sci. USA* **110**, E4474–E4481 (2013).
25. Finley, K. D., Taylor, B. J., Milstein, M. & McKeown, M. *dissatisfaction*, a gene involved in sex-specific behavior and neural development of *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **94**, 913–918 (1997).
26. Robinett, C. C., Vaughan, A. G., Knapp, J. M. & Baker, B. S. Sex and the single cell. II. There is a time and place for sex. *PLoS Biol.* **8**, e1000365 (2010).
27. Markle, J. G. *et al.* Sex differences in the gut microbiome drive hormone-dependent regulation of autoimmunity. *Science* **339**, 1084–1088 (2013).
28. Steegenga, W. T. *et al.* Sexually dimorphic characteristics of the small intestine and colon of prepubescent C57BL/6 mice. *Biol. Sex Diff.* **5**, 11 (2014).
29. Cook, M. B. *et al.* Sex disparities in cancer incidence by period and age. *Cancer Epidemiol. Biomarkers Prev.* **18**, 1174–1182 (2009).
30. Siudeja, K. *et al.* Frequent somatic mutation in adult intestinal stem cells drives neoplasia and genetic mosaicism during aging. *Cell Stem Cell* **17**, 663–674 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank K. Vanezis and L. Game for technical assistance, and C. Gonzalez, T. Carroll and D. Perea for discussions. B. Baker, A. Bardin, A. Baena-Lopez, S. Bray, A. Casali, T. Cline, S. Cohen, D. Eberl, S. Goodwin, G. Jefferis, L. Jones, R. Niwa, B. Prud'homme, I. Salecker, L. Sanchez, C. Thummel, J. Treisman, M. Vidal, D. Yamamoto and L. Zwiebel shared reagents. D. Hadjieconomou, J. Jacobson, G. King and E. Parra-Peralbo provided comments on the manuscript. This work was funded by an ERC Starting Grant to I.M.-A. (ERCStG 310411) and MRC intramural funding. B.H. holds an EMBO long-term fellowship and I.M.-A. is a member of, and is supported by, the EMBO Young Investigator Programme.

Author Contributions B.H. and I.M.-A. designed and conceived the study. S.K. performed statistical analyses of RNA-seq data, B.H. performed all other experiments in the study and analysed data. I.M.-A. analysed data and wrote the manuscript, with contributions from B.H.

Author Information RNA-seq transcriptional profiling of virgin adult midguts has been deposited at GEO under accession number GSE74775. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to I.M.-A. (i.miguel-aliaga@imperial.ac.uk) or B.H. (bruno.hudry@imperial.ac.uk).

METHODS

Fly husbandry. Fly stocks were reared on a standard cornmeal/agar diet (6.65% cornmeal, 7.15% dextrose, 5% yeast, 0.66% agar supplemented with 2.2% nipagin and 3.4 ml l⁻¹ propionic acid). All experimental flies were kept in incubators at 25 °C, 65% humidity and on a 12 h light/dark cycle, except for those containing *tub-Gal80^{TS}* transgenes, which were set up at 18 °C (restrictive temperature) and transferred to 29 °C (permissive temperature) at the time when *Gal4* induction was required. This depended on the specific experimental requirements but typically, for loss-of-function (RNAi) or gain-of-function (UAS) experiments, flies were raised and aged as adults for 3 days at 18 °C, were then shifted to 29 °C to induce transgene expression, and adult midguts were dissected after 10–20 days (as indicated in each figure panel). Flies were transferred to fresh vials every 3 days and fly density was kept to a maximum of 15 flies per vial. Virgin flies were used for all experiments unless indicated otherwise.

For mutant ISC clonal analyses (MARCM clones), 3-day-old adults (raised and aged at 25 °C) were heat-shocked for 1 h at 37 °C to induce clones, and were then kept at 25 °C (or 29 °C for MARCM RNAi clones) until dissection (10 days, 15 days or 4 weeks thereafter as indicated in each figure panel). Flies were transferred to fresh vials every 3 days.

For damage-induced regeneration assays, virgin flies were collected over 72 h at 18 °C, and were then shifted to 29 °C for 7 days on standard media. Flies were then transferred in an empty vial containing a piece of 3.75 cm × 2.5 cm paper. 500 ml of 5% sucrose solution (control) or 5% sucrose + 3 dextran sulphate sodium (DSS) solution was used to wet the paper, used as feeding substrate. Flies were transferred to new vial with fresh feeding paper every day for 3 days prior dissection.

For fecundity experiments, females were raised and aged as virgins for 3 days at 18 °C, and were then shifted to 29 °C to induce transgene expression for 10 days. Females were then mated overnight to OregonR males (10 males, 10 females per vial). Males were then removed, and single female flies were transferred individual vials every 24 h for a 3-day period. Eggs were counted from the vacated vials.

Fly stocks. *UAS transgenes.* *UAS-Sxl^{RNAi}* (VDRC: GD 3131), *UAS-Sxl* (generated in ref. 31), *UAS-pon.GFP* (BDSC: 42741, generated in ref. 32), *UAS-GFP.E2fl.1⁻²³⁰*, *UAS-mRFP1.NLS.CycB.1⁻²⁶⁶* (BDSC: 55121, generated in ref. 33), *UAS-GFP* (BDSC: 35786), *UAS-Dicer2* (VDRC: 60007), *UAS-tra^{RNAi}* (BDSC: 28512, TRiP.JF03132), *UAS-tra^F* (BDSC: 4590, generated in ref. 34), *UAS-msl-2^{RNAi}* (VDRC: GD 29356), *UAS-Sxl^{RNAi} 2* (VDRC: KK 109221), *UAS-msl-2^{RNAi} 2* (BDSC: 35390, TRiP.GL00309), *UAS-msl-2^{HA}* (this study), *UAS-dsx^{RNAi}* (BDSC: 26716, TRiP.JF02256), *UAS-dsx^F* (BDSC: 44223, generated in ref. 35), *UAS-Arr1* (generated in ref. 36), *UAS-rdo* (generated in ref. 17), *UAS-GstE4* (generated in ref. 37), *UAS-Notch^{RNAi}* (VDRC: KK 100002), *UAS-Sxl^{RNAi} 3* (BDSC: 34393, TRiP.HMS00609), *UAS-tra^{RNAi} 2* (VDRC: GD 2560), *UAS-tra^{RNAi} 3* (NIG: 16724-2R2), *UAS-fru^{RNAi}* (BDSC: 31593, TRiP.JF01182), *UAS-Notch^{RNAi} 2* (VDRC: GD 27229), *UAS-Hairless* (generated in ref. 38), *UAS-N^{intra}* (generated in ref. 39), *UAS-sspitzi* (gift from J. Treisman, generated in ref. 40), *UAS-Ras^{V12}*; *FRT82B* line (generated in ref. 41).

Mutants. *tra¹, FRT2A, fru^{P1}-Gal4* chromosome (generated in ref. 42), *FRT82B, dsx¹* chromosome (generated in ref. 43), *dsx^A, Df(3R)^{Exel6179}* chromosome (generated in ref. 44), *Df(3R)dsx¹¹* (BDSC: 1865, generated in ref. 45), *In(3R)dsx²³* (BDSC: 1849, generated in ref. 46), *FRT82B, dsx^A, fru^{P1-LexA}* chromosome (gift from B. Baker, generated in ref. 47), *dsx^A, fru^{P1-LexA}* chromosome (gift from B. Prud'homme, generated in ref. 48), *tra^{KO}* (this study), *Df(3L)^{st-17}* (BDSC: 5416, generated in ref. 49), *tra¹* (BDSC: 675, generated in ref. 50), *tra^{2B}* (BDSC: 25137, generated in ref. 51), *Df(2R)^{trix}* (BDSC: 1896, generated in ref. 52), *tra^{2S1}* (BDSC: 2413, generated in ref. 52), *dsx^D* (BDSC: 840, generated in ref. 53), *fru^F* and *fru^M* (gift from G. Jefferis, generated in ref. 47), *fru⁴⁻⁴⁰* (gift from G. Jefferis, generated in ref. 54), *FRT82B, Df^{RevF10}* chromosome (gift from S. Bray, generated in ref. 55), *Su(H)^{Δ47}*, *FRT40A* chromosome (gift from S. Bray, generated in ref. 56), *FRT82B Apc2^{N175K}*, *Apc^{Q8}* chromosome and *hs-flp; UAS-Ras^{V12}*; *FRT82B Apc2^{N175K}, Apc^{Q8}* line (both generated in ref. 41).

Reporters and Gal4 drivers. *Su(H)-LacZ* (generated in ref. 57), *esg-GFP^{P01986}* (gift from L. Jones, generated in ref. 88), *esg-LacZ^{k00606}* (BDSC: 10359, generated in ref. 89), *dsx-Gal4* (generated in ref. 58), *fru^{P1}-Gal4* (gift from G. Jefferis, generated in ref. 59), *esg-Gal4^{NP7397}*, *UAS-GFP; Tub-Gal80^{TS}* chromosome (gift from J. de Navascués), *Su(H)^{GBE}-Gal80* (generated in ref. 60), *Su(H)^{GBE}-Gal4* (generated in ref. 61), *mex1-Gal4* (generated in ref. 62), *pros^{V1}-Gal4* (generated in ref. 63), *Mvl-Gal4^{NP2375}* (Kyoto: 104178), *dsx^{Δ2}-Gal4* (generated in ref. 64), *vm-Gal4* (BDSC: 48547, GMR13B09-GAL4), *btl-Gal4* (generated in ref. 65), *nSyb-Gal4* (BDSC: 51941, generated in ref. 90), *elav-Gal4* (BDSC: 8765, generated in ref. 91), *nSyb-Gal80* (gift from J. Simpson), *stripe-Gal4* (BDSC: 26663, generated in ref. 92). **MARCM stocks.** *FRT40A: w, hs-flp, Tub-Gal4, UAS-GFP; tub-Gal80, FRT40A* (gift from J. de Navascués), *FRT40A* (gift from J. de Navascués, generated in ref. 93),

FRT2A: y, w, hs-flp; Tub-Gal4, UAS-GFP; Tub-Gal80, FRT2A (gift from I. Salecker), *FRT2A* (BDSC: 1997, generated in ref. 94), *FRT82B: w, hs-flp; UAS-mCD8GFP, Tub-Gal4; FRT82B, Tub-Gal80* (gift from M. Vidal), *FRT82B* (BDSC: 7369, generated in ref. 93).

Generation of *tra^{KO}* and *tra^{LacZ}* transgenic lines. To generate the transgenic reporter of *tra* promoter activity *tra^{LacZ}*, the upstream inter-genic region between *spd-2* and *tra* was cloned using the following primer pairs: 5'-AAAATCTAGAACTAATAAGTATATGAG-3' and 5'-AAAAGGTA CCCGAAAATGCTGGAAATTAAATGATGC-3'. PCR was performed with Q5 high-fidelity polymerase from New England Biolabs (M0491S). The PCR product was digested with *Xba*I and *Kpn*I, and was then cloned into the *pH-Pelican attb* vector (ref. 66, gift from C. Thummel). The construct was sequence-verified and a transgenic line was established through *ΦC-31* integrase mediated transformation (Bestgene, *attP* site: *attP40*).

The new amorphic allele of *tra*, *tra^{KO}*, was generated using the accelerated homologous recombination method recently developed by Baena-Lopez⁶⁷. The 5' (4,302 nt) and 3' (2,928 nt) homology arms were produced from a *tra* BAC (BAC-PAC resources: CH322-140P01) with the following primer pairs respectively: 5'-AAAAGCGCCGCCATTCTACTCTTGAATTGGCTAGC-3'/5'-AAAAGTA CCATGATGCACTTCTCAGTGTGA-3' and 5'-AAAAGCGCGCGCAAG AGAACATCATGG-3'/5'-AAAAGCGCGCCATTGTCGACACAATCAAATCG-3'. PCRs were performed with Q5 high-fidelity polymerase from New England Biolabs (M0491S). The PCR products were digested with *Not*I/*Kpn*I or *Asci* respectively, and were then cloned into the *pTV^{Cherry}* vector to generate *pTV^{Cherry}[tra]*. The vector was sequence-verified and inserted into random genomic locations by P-element-mediated transformation (Bestgene). Transformants (not necessarily mapped or homozygosed) were crossed to *hs-FLP*, *hs-SceI* flies (DBSC: 25679) and the resulting larvae were heat-shocked at 48, 72, 96 and 120 h after egg laying for 1 h at 37 °C. Approximately 200 adult females with mottled eyes (indicating the presence of *pTV^{Cherry}[tra]* and the transgenes carrying *hs-FLP* and *hs-SceI*) were crossed in pools of 15 to *ubiquitin-Gal4[3xP3-GFP]* males and the progeny was screened for the presence of red-eyed flies. The *ubiquitin-Gal4[3xP3-GFP]* transgene was subsequently removed by selecting against the presence of GFP in the ocelli. The generated deletion removed 342 nucleotides starting 13 nucleotides upstream of the transcription start site.

Generation of the *UAS-msl-2^{HA}* line. To overexpress *msl-2*, a transgenic *UAS* line was generated from *msl-2* cDNA adding an HA tag at the C-terminal (BDGP Gold cDNAs, clone ID: GH22488) with the following primer pair: 5'-AAAAGATC TATGGCTCAGACGGCATACTTG and 5'-AAAATCTAGATTAAGCGTAATC TGGAACATCGTATGGTACAAGTCATCCGAGCCGCAC-3'. PCR was performed with Q5 high-fidelity polymerase from New England Biolabs (M0491S). The PCR product was digested with *Bgl*II and *Xba*I before cloning into the *pUASTattb* vector⁶⁸. The construct was sequence-verified and a transgenic line was established through *ΦC-31* integrase mediated transformation (Bestgene, *attP* site ZH-86Fb, DBSC: 24749).

Immunohistochemistry. Intact guts were fixed at room temperature for 20 min in PBS, 3.7% formaldehyde. All subsequent incubations were done in PBS, 4% horse serum, 0.2% Triton X-100 at 4 °C following standard protocols.

The following primary antibodies were used: mouse anti-Sxl (M114, DSHB Hydridoma) 1:50, mouse anti-Sxl (M18, DSHB Hydridoma) 1:50, goat anti-Msl-2 (dC-20, sc-32459, Santa Cruz Biotechnology) 1:50, chicken anti-beta galactosidase (ab9361, Abcam) 1:200, rabbit anti-phospho-histone H3 Ser10 (9701L, Cell Signalling Technology) 1:500, rabbit anti-fru^M (Male-2, generated in ref. 69) 1:500, mouse anti-Dsx^{DBD} (DSHB Hydridoma) 1:100, mouse anti-Pros (MR1A, DSHB Hydridoma) 1:50, goat anti-ac-Histone H4 Lys16 (sc-8662, Santa Cruz Biotechnology) 1:500, rat anti-HA (11867423001, Roche) 1:500, mouse anti-GFP (11814460001, Roche) 1:1,000, mouse anti-Pdm1 (kind gift of Steve Cohen, generated in ref. 70) 1:20. Fluorescent secondary antibodies (FITC-, Cy3- and Cy5-conjugated) were obtained from Jackson Immunoresearch. Vectashield with DAPI (Vector Labs) was used to stain DNA.

Cell, clone and midgut length quantifications. Mitotic indices were quantified by counting phospho-Histone H3-positive cells in >10 midguts per genotype, time point, and/or condition (for example, male/female), and are displayed as means ± standard error of the mean (s.e.m.). For posterior midgut cell counts, a midgut portion immediately anterior to the junction with the tubules and hindgut was imaged at 20× magnification. Cells were counted using ImageJ in areas of identical size across all genotypes to control for size differences. Threshold was adjusted for the GFP channel (ImageJ function: Image > Adjust > Threshold) to subtract background, and the percentage of area above the threshold was considered (ImageJ function: analyse particles). Data was collected from at least 10 midguts per genotype and/or condition, and is displayed as mean of percentage of GFP-positive area ± s.e.m.

MARCM clones were quantified as number of cells per clone (when the GFP reporter was expressed in the nucleus) or by the size of the GFP area/clone (in arbitrary units, when the reporter was membrane-tagged GFP).

To measure midgut length, guts were dissected and were then straightened on polylysine-coated slides. After imaging, a line was drawn between the most anterior point of the proventriculus and the midgut–hindgut junction using ImageJ. The number of pixels contained in the line was used as a proxy for midgut length. For display purposes, the value obtained for a control female midgut selected at random was used as reference to normalize all the other values, which are shown as percentage of control female length value.

Functional screen of *tra* targets. To test the functional significance of *tra* targets, only those with female/*tra* null mutant female fold differences in transcript abundance $>\pm 2$ that were also under *tra* control in adult intestinal progenitors were selected for functional validation. The contribution of these *tra* targets to sex differences in adult ISC proliferation was investigated using publicly available RNAi and UAS lines (see below for details of the RNAi lines, and Methods section ‘Fly stocks’ for details of the UAS lines) expressed from *esg-Gal4^{ts}*. Results obtained with 11 out of the 16 lines belonging to the VDRC KK collection were not considered and are not shown in Extended Data Fig. 7 because their expression in adult ISCs from *esg-Gal4^{ts}* resulted in a recurrent ISC differentiation phenotype in both males and females. We attribute this effect to the previously reported dominant *Gal4*-dependent toxicity issue with this VDRC KK collection⁷¹. The genes from the GD library were *ldgf1* (transformant ID 12416); *Spn88Eb* (transformant ID 28425); *CG8008* (transformant ID 4159); *CG4500* (transformant ID 34852); *Arr1* (transformant ID 22196); *Spn47C* (transformant ID 25534); *Cyp12d1-p* (transformant ID 49269); *TwdlE* (transformant ID 24867); *Cyp12d1-d* (transformant ID 50507); *GstE2* (transformant ID 32945); *GstE4* (transformant ID 20472); *lectin-28C* (transformant ID 45634); *AANATL2* (transformant ID 44677); *CG5348* (transformant ID 1698); *CG15128* (transformant ID 19302) and *Ir10a* (transformant ID 45403). The genes from the KK library were *CG17470* (transformant ID 100414); *Klp54D* (transformant ID 100140); *CG4500* (transformant ID 106260); *Arr1* (transformant ID 109860); *rdo* (transformant ID 107213); *Spn47C* (transformant ID 105933); *Cyp12d1-p* (transformant ID 109256); *TwdlE* (transformant ID 107483); *Cyp12d1-d* (transformant ID 109248); *GstE4* (transformant ID 100986); *lectin-28C* (transformant ID 104290); *AANATL2* (transformant ID 102802); *CG5348* (transformant ID 106565); *CG15128* (transformant ID 100238); *CG15236* (transformant ID 101045) and *Ir10a* (transformant ID 100181).

Statistics and data presentation. All statistical analyses were carried out in the R environment⁷². Comparisons between two genotypes/conditions were analysed with the Mann–Whitney–Wilcoxon rank sum test (R function *wilcox.test*). All graphs were generated using Adobe Illustrator. All confocal and bright field images belonging to the same experiment and displayed together in our figure were acquired using the exact same settings. For visualization purposes, level and channel adjustments were applied using ImageJ to the confocal images shown in the figure panels (the same correction was applied to all images belonging to the same experiment), but all quantitative analyses were carried out on unadjusted raw images or maximum projections. In all figures, *n* denotes the number of midguts, ISCs/EBs, mitoses or clones that were analysed for each genotype. Values are presented as average \pm standard error of the mean (s.e.m.), *P* values from Mann–Whitney–Wilcoxon test (non-significant (NS): *P* > 0.05; **P* < 0.05 > *P* > 0.01; ***P* < 0.01 > *P* > 0.001; ****P* < 0.001). Lines and asterisks highlighting significant comparisons across sexes are displayed in red, whereas those highlighting significant comparisons within same-sex data sets are displayed in black.

The number of *Drosophila* animals was not limiting, so no statistical methods were used to predetermine sample size. The experiments were not randomized.

RNA-seq experiments. **RNA extraction.** RNA from 30 pooled dissected midguts was extracted using TRIzol (Invitrogen), and 3 such samples were used for each sex and genotype. RNA-seq libraries were prepared from 500ng of total RNA using the Illumina Truseq mRNA stranded library prep kit (Illumina, San Diego, USA) according to the manufacturer’s protocol. Library quality was checked on a Bioanalyser HS DNA chip and concentrations were estimated by Qubit measurement. Libraries were pooled in equimolar quantities and sequenced on a HiSeq2500 using paired end 100bp reads. At least 30 million reads passing filter were achieved per sample. After demultiplexing, raw RNASeq reads were aligned with Tophat splice junction mapper⁷³, version 2.0.11 against Ensembl *Drosophila* genome reference sequence assembly (dm3) and transcript annotations.

Differential gene expression analysis. For differential gene expression analysis, gene-based read counts were then obtained using HTSeq count module (version 0.5.3p9)⁷⁴. Differential expression analysis was performed on the counts data using DESeq2 Bioconductor package⁷⁵. The analysis was run with the default parameters. DESeq2 package uses negative binomial model to model read counts and

then performs statistical tests for differential expression of genes. Raw *P* values were then adjusted for multiple testing with the Benjamini–Hochberg procedure. GO enrichment analysis was performed using FlyMine v40.1.b. In Fig. 3a, b and Extended Data Fig. 1g, each column corresponds to one of three different replicates (30 midguts each) for each sex, and transcript abundance for each gene was normalized to a scale of -2 (white) to $+2$ (grey).

Isoform expression analysis. Reads were first aligned to the transcript sets using Bowtie software (version 1.0.0)⁷⁶. Transcript annotations for cDNA and non-coding RNAs were obtained from Ensembl (release 75). Isoform abundances were then calculated using mmseq R package (version 1.0.8)⁷⁷. Differential expression analysis was performed on these abundances using the DESeq2 Bioconductor package (version 1.42.1)⁷⁸. Isoforms not expressed in any of the samples were not included in the differential expression analysis. Raw *P* values were adjusted for multiple testing with the Benjamini–Hochberg procedure. Initial isoform analysis identified differential expression for 1,379 isoforms between male and female midguts (*P* < 0.05 cutoff), which, following subtraction of those corresponding to single-isoform genes, yielded 704 isoforms (see main text for subsequent analysis).

RNA-seq data displays. For the volcano plots in Extended Data Fig. 1, \log_2 fold change values were plotted against \log_{10} of adjusted *P* values. Genes and isoforms significantly upregulated in males are coloured blue, while those significantly upregulated in females are coloured red. Selected genes and isoforms are further highlighted as empty circles. For the scatter charts with quadrants in Extended Data Fig. 7, \log_2 fold change values (between control and *tra* mutant females) were plotted against \log_2 fold change values (between *tra* mutant females and *tra* mutant females and female with adult-specific, progenitor-specific *tra* expression). Genes and isoforms significantly repressed by *tra* are coloured blue, whereas those significantly activated by *tra* are coloured red. Selected genes and isoforms are further highlighted as empty circles. Volcano plots and scatter charts were generated using Adobe Illustrator. To generate the heat maps, a matrix containing the relative values of gene/isoform expression was built. Transcript abundance for each gene/isoform was normalized to a scale ranging from -2 to $+2$. A hierarchical clustering algorithm (with Euclidian distance and average linking) was applied to the matrix using the MeV software suite⁷⁹. Area-proportional Venn diagrams were generated using BioVenn (<http://www.cmbi.ru.nl/cdd/biovenn/>).

Real time qRT-PCRs. RNAs were extracted from 30 dissected midguts using TRIzol (Invitrogen). RNAs were cleaned using RNAeasy mini Kit (QIAGEN), and cDNAs were synthesized using the iScript cDNA synthesis kit (Bio-Rad) from 300ng of total RNAs. Quantitative PCRs were performed by mixing cDNA samples (5ng) with iTaq Universal SYBR Green Supermix (Bio-Rad, #172-5124) and the relevant primers in a 96-well plate. Expression values were normalized to *eIF4G*. For each gene/isoform, at least 3 independent biological replicates were used, and 2 technical replicates were performed. See Supplementary Information for a list of all primers used.

- Ruiz, M. F., Sarno, F., Zorrilla, S., Rivas, G. & Sanchez, L. Biochemical and functional analysis of *Drosophila*-*sciara* chimeric sex-lethal proteins. *PLoS ONE* **8**, e65171 (2013).
- Lu, B., Ackerman, L., Jan, L. Y. & Jan, Y. N. Modes of protein movement that lead to the asymmetric localization of partner of Numb during *Drosophila* neuroblast division. *Mol. Cell* **4**, 883–891 (1999).
- Zielke, N. *et al.* Fly-FUCCI: A versatile tool for studying cell proliferation in complex tissues. *Cell Rep.* **7**, 588–598 (2014).
- Ferveur, J. F., Stortkuhl, K. F., Stocker, R. F. & Greenspan, R. J. Genetic feminization of brain structures and changed sexual orientation in male *Drosophila*. *Science* **267**, 902–905 (1995).
- Lee, G., Hall, J. C. & Park, J. H. Doublesex gene expression in the central nervous system of *Drosophila melanogaster*. *J. Neurogenet.* **16**, 229–248 (2002).
- Merrill, C. E., Sherertz, T. M., Walker, W. B. & Zwiebel, L. J. Odorant-specific requirements for arrestin function in *Drosophila* olfaction. *J. Neurobiol.* **63**, 15–28 (2005).
- Dabor, P. J. *et al.* Evaluating the insecticide resistance potential of eight *Drosophila melanogaster* cytochrome P450 genes by transgenic over-expression. *Insect Biochem. Mol. Biol.* **37**, 512–519 (2007).
- Bardin, A. J., Perdigoto, C. N., Southall, T. D., Brand, A. H. & Schweigert, F. Transcriptional control of stem cell maintenance in the *Drosophila* intestine. *Development* **137**, 705–714 (2010).
- de Celis, J. F. & Bray, S. Feed-back mechanisms affecting Notch activation at the dorsoventral boundary in the *Drosophila* wing. *Development* **124**, 3241–3251 (1997).
- Schweitzer, R., Shaharabany, M., Seger, R. & Shilo, B. Z. Secreted Spitz triggers the DER signaling pathway and is a limiting component in embryonic ventral ectoderm determination. *Genes Dev.* **9**, 1518–1529 (1995).
- Martorell, O. *et al.* Conserved mechanisms of tumorigenesis in the *Drosophila* adult midgut. *PLoS ONE* **9**, e88413 (2014).
- Kohl, J., Ostrovsky, A. D., Frechter, S. & Jefferis, G. S. A bidirectional circuit switch reroutes pheromone signals in male and female brains. *Cell* **155**, 1610–1623 (2013).

43. Mellert, D. J., Robinett, C. C. & Baker, B. S. *doublesex* functions early and late in gustatory sense organ development. *PLoS ONE* **7**, e51489 (2012).

44. Cande, J., Stern, D. L., Morita, T., Prud'homme, B. & Gompel, N. Looking under the lamp post: neither *fruitless* nor *doublesex* has evolved to generate divergent male courtship in *Drosophila*. *Cell Rep.* **8**, 363–370 (2014).

45. Baker, B. S., Hoff, G., Kaufman, T. C., Wolfner, M. F. & Hazelrigg, T. The *doublesex* locus of *Drosophila melanogaster* and its flanking regions: a cytogenetic analysis. *Genetics* **127**, 125–138 (1991).

46. Duncan, I. W. & Kaufman, T. C. Cytogenetic analysis of chromosome 3 in *Drosophila melanogaster*: mapping of the proximal portion of the right arm. *Genetics* **80**, 733–752 (1975).

47. Demir, E. & Dickson, B. J. *fruitless* splicing specifies male courtship behavior in *Drosophila*. *Cell* **121**, 785–794 (2005).

48. Mellert, D. J., Knapp, J. M., Manoli, D. S., Meissner, G. W. & Baker, B. S. Midline crossing by gustatory receptor neuron axons is regulated by *fruitless*, *doublesex* and the Roundabout receptors. *Development* **137**, 323–332 (2010).

49. Belote, J. M., Hoffmann, F. M., McKeown, M., Chorsky, R. L. & Baker, B. S. Cytogenetic analysis of chromosome region 73AD of *Drosophila melanogaster*. *Genetics* **125**, 783–793 (1990).

50. Sturtevant, A. H. A gene in *Drosophila melanogaster* that transforms females into males. *Genetics* **30**, 297–299 (1945).

51. Mattox, W. & Baker, B. S. Autoregulation of the splicing of transcripts from the *transformer-2* gene of *Drosophila*. *Genes Dev.* **5**, 786–796 (1991).

52. Belote, J. M. & Baker, B. S. Sexual behavior: its genetic control during development and adulthood in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **84**, 8026–8030 (1987).

53. Gowen, J. W. & Fung, S. T. C. Determination of sex through genes in a major sex locus in *Drosophila melanogaster*. *Heredity* **11**, 397–402 (1957).

54. Anand, A. *et al.* Molecular genetic dissection of the sex-specific and vital functions of the *Drosophila melanogaster* sex determination gene *fruitless*. *Genetics* **158**, 1569–1595 (2001).

55. Haenlin, M., Kramatschek, B. & Campos-Ortega, J. A. The pattern of transcription of the neurogenic gene *Delta* of *Drosophila melanogaster*. *Development* **110**, 905–914 (1990).

56. Morel, V. & Schweigert, F. Repression by suppressor of *hairless* and activation by Notch are required to define a single row of *single-minded* expressing cells in the *Drosophila* embryo. *Genes Dev.* **14**, 377–388 (2000).

57. Furriols, M. & Bray, S. A model Notch response element detects Suppressor of *Hairless*-dependent molecular switch. *Curr. Biol.* **11**, 60–64 (2001).

58. Rideout, E. J., Dornan, A. J., Neville, M. C., Eadie, S. & Goodwin, S. F. Control of sexual differentiation and behavior by the *doublesex* gene in *Drosophila melanogaster*. *Nature Neurosci.* **13**, 458–466 (2010).

59. Manoli, D. S. *et al.* Male-specific *fruitless* specifies the neural substrates of *Drosophila* courtship behaviour. *Nature* **436**, 395–400 (2005).

60. Wang, L., Zeng, X., Ryoo, H. D. & Jasper, H. Integration of UPR^{ER} and oxidative stress signaling in the control of intestinal stem cell proliferation. *PLoS Genet.* **10**, e1004568 (2014).

61. Zeng, X., Chauhan, C. & Hou, S. X. Characterization of midgut stem cell- and enteroblast-specific Gal4 lines in *Drosophila*. *Genesis* **48**, 607–611 (2010).

62. Phillips, M. D. & Thomas, G. H. Brush border spectrin is required for early endosome recycling in *Drosophila*. *J. Cell Sci.* **119**, 1361–1370 (2006).

63. Balakireva, M., Stocker, R. F., Gendre, N. & Ferveur, J. F. *Voila*, a new *Drosophila* courtship variant that affects the nervous system: behavioral, neural, and genetic characterization. *J. Neurosci.* **18**, 4335–4343 (1998).

64. Pan, Y., Robinett, C. C. & Baker, B. S. Turning males on: activation of male courtship behavior in *Drosophila melanogaster*. *PLoS ONE* **6**, e21144 (2011).

65. Shiga, Y., Tanaka-Matakatsu, M. & Hayashi, S. A nuclear GFP/β-galactosidase fusion protein as a marker for morphogenesis in living *Drosophila*. *Dev. Growth Differ.* **38**, 99–106 (1996).

66. Misra, J. R., Horner, M. A., Lam, G. & Thummel, C. S. Transcriptional regulation of xenobiotic detoxification in *Drosophila*. *Genes Dev.* **25**, 1796–1806 (2011).

67. Baena-Lopez, L. A., Alexandre, C., Mitchell, A., Pasakarnis, L. & Vincent, J. P. Accelerated homologous recombination and subsequent genome modification in *Drosophila*. *Development* **140**, 4818–4825 (2013).

68. Bischof, J., Maeda, R. K., Hediger, M., Karch, F. & Basler, K. An optimized transgenesis system for *Drosophila* using germ-line-specific ϕ C31 integrases. *Proc. Natl. Acad. Sci. USA* **104**, 3312–3317 (2007).

69. Kimura, K., Ote, M., Tazawa, T. & Yamamoto, D. *Fruitless* specifies sexually dimorphic neural circuitry in the *Drosophila* brain. *Nature* **438**, 229–233 (2005).

70. Yeo, S. L. *et al.* On the functional overlap between two *Drosophila* POU homeo domain genes and the cell fate specification of a CNS neural precursor. *Genes Dev.* **9**, 1223–1236 (1995).

71. Green, E. W., Fedele, G., Giorgini, F. & Kyriacou, C. P. A *Drosophila* RNAi collection is subject to dominant phenotypic effects. *Nature Methods* **11**, 222–223 (2014).

72. R Development Core Team. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2014).

73. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).

74. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

75. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).

76. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).

77. Turro, E. *et al.* Haplotype and isoform specific expression estimation using multi-mapping RNA-seq reads. *Genome Biol.* **12**, R13 (2011).

78. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

79. Saeed, A. I. *et al.* TM4 microarray software suite. *Methods Enzymol.* **411**, 134–193 (2006).

80. Jiang, H. & Edgar, B. A. Intestinal stem cell function in *Drosophila* and mice. *Curr. Opin. Genet. Dev.* **22**, 354–360 (2012).

81. Lemaitre, B. & Miguel-Aliaga, I. The digestive tract of *Drosophila melanogaster*. *Annu. Rev. Genet.* **47**, 377–404 (2013).

82. Kelley, R. L. *et al.* Expression of *msl-2* causes assembly of dosage compensation regulators on the X chromosomes and female lethality in *Drosophila*. *Cell* **81**, 867–877 (1995).

83. Lee, T. & Luo, L. Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. *Neuron* **22**, 451–461 (1999).

84. Jiang, H. *et al.* Cytokine/Jak/Stat signaling mediates regeneration and homeostasis in the *Drosophila* midgut. *Cell* **137**, 1343–1355 (2009).

85. Patel, P. H., Dutta, D. & Edgar, B. A. Niche appropriation by *Drosophila* intestinal stem cell tumours. *Nature Cell Biol.* **17**, 1182–1192 (2015).

86. Buchon, N. *et al.* Morphological and molecular characterization of adult midgut compartmentalization in *Drosophila*. *Cell Rep.* **3**, 1725–1738 (2013).

87. Micchelli, C. A. & Perrimon, N. Evidence that stem cells reside in the adult *Drosophila* midgut epithelium. *Nature* **439**, 475–479 (2006).

88. Kelso, R. J. *et al.* Flytrap, a database documenting a GFP protein-trap insertion screen in *Drosophila melanogaster*. *Nucleic Acids Res.* **32**, D418–D420 (2004).

89. Spradling, A. C. *et al.* The Berkeley *Drosophila* genome project gene disruption project. Single P-element insertions mutating 25% of vital *Drosophila* genes. *Genetics* **153**, 135–177 (1999).

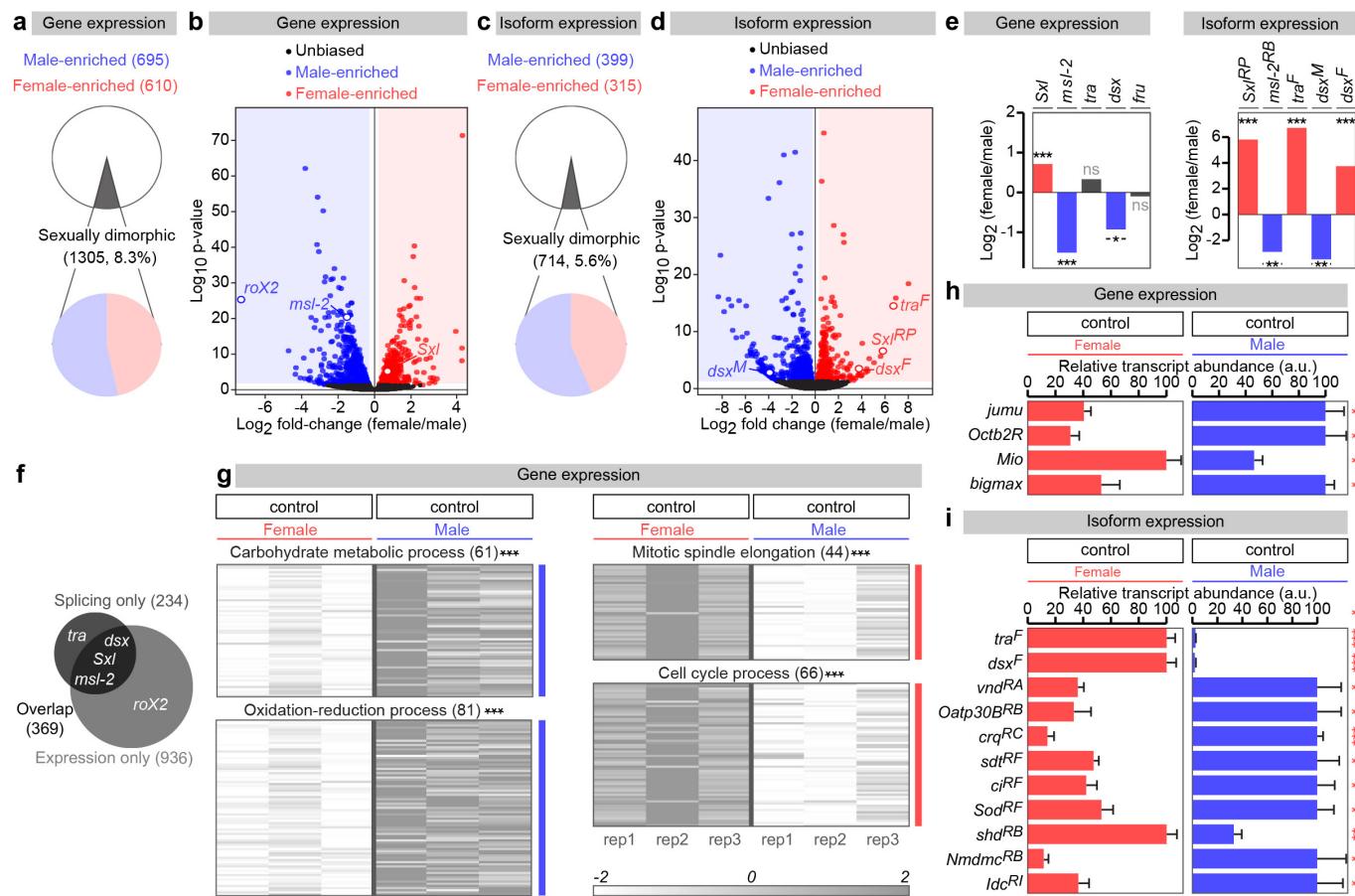
90. Riabinina, O. *et al.* Improved and expanded Q-system reagents for genetic manipulations. *Nature Methods* **12**, 219–222 (2015).

91. Luo, L., Liao, Y. J., Jan, L. Y. & Jan, Y. N. Distinct morphogenetic functions of similar small GTPases: *Drosophila* Drac1 is involved in axonal outgrowth and myoblast fusion. *Genes Dev.* **8**, 1787–1802 (1994).

92. Ghazi, A., Anant, S. & VijayRaghavan, K. Apterous mediates development of direct flight muscles autonomously and indirect flight muscles through epidermal cues. *Development* **127**, 5309–5318 (2000).

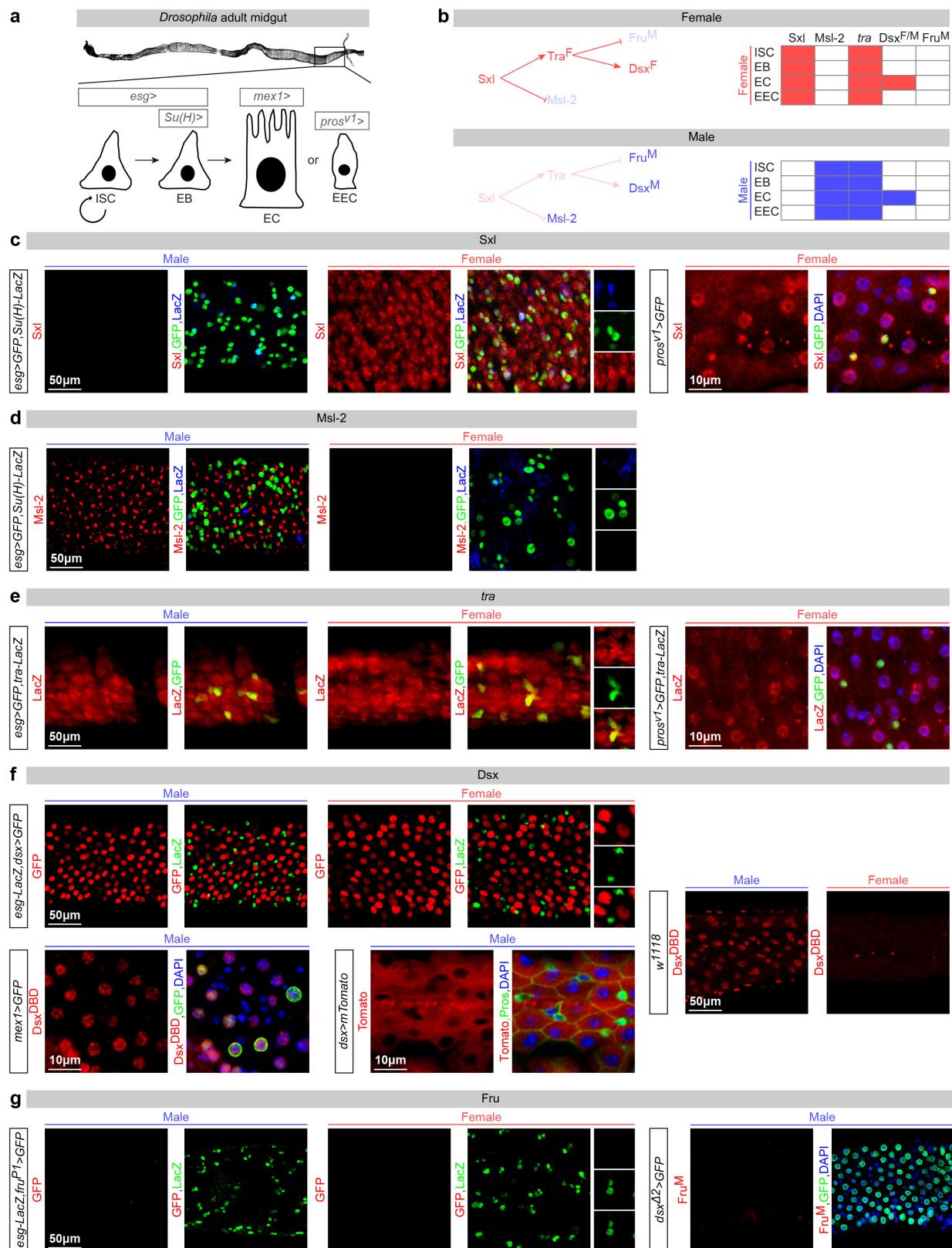
93. Xu, T. & Rubin, G. M. Analysis of genetic mosaics in developing and adult *Drosophila* tissues. *Development* **117**, 1223–1237 (1993).

94. Golic, K. G. Site-specific recombination between homologous chromosomes in *Drosophila*. *Science* **252**, 958–961 (1991).



Extended Data Figure 1 | Sexually dimorphic transcription and splicing in the adult midgut. **a**, Number and percentage of genes with sexually dimorphic gene expression, as revealed by RNA-seq transcriptional profiling of virgin male and female dissected midguts ($P < 0.05$ cutoff). **b**, Volcano plot displaying all genes with detectable midgut expression. Female/male ratio of gene expression is shown on the *x* axis (in log₂ scale) and significance is displayed on the *y* axis as the negative logarithm (log₁₀ scale) of the adjusted P value. Genes with significantly upregulated ($P < 0.05$ cutoff) expression in males and females are coloured in blue and red, respectively. Other genes are displayed in black. Genes with known sex-specific transcription are displayed as red (female-enriched) or blue (male-enriched) open circles. **c, d**, Comparable analyses for sex-biased isoforms belonging to genes with multiple transcripts. We identified 714 sex-biased isoforms belonging to a total of 603 genes. Isoforms resulting from known sex-specific alternative splicing are displayed as in panel **b**. **e**, Female/male ratios of overall transcript abundance (left graph) and abundance of sex-biased isoforms (right graph) for the members of the *Drosophila* sex determination pathway as revealed by RNA-seq analysis of

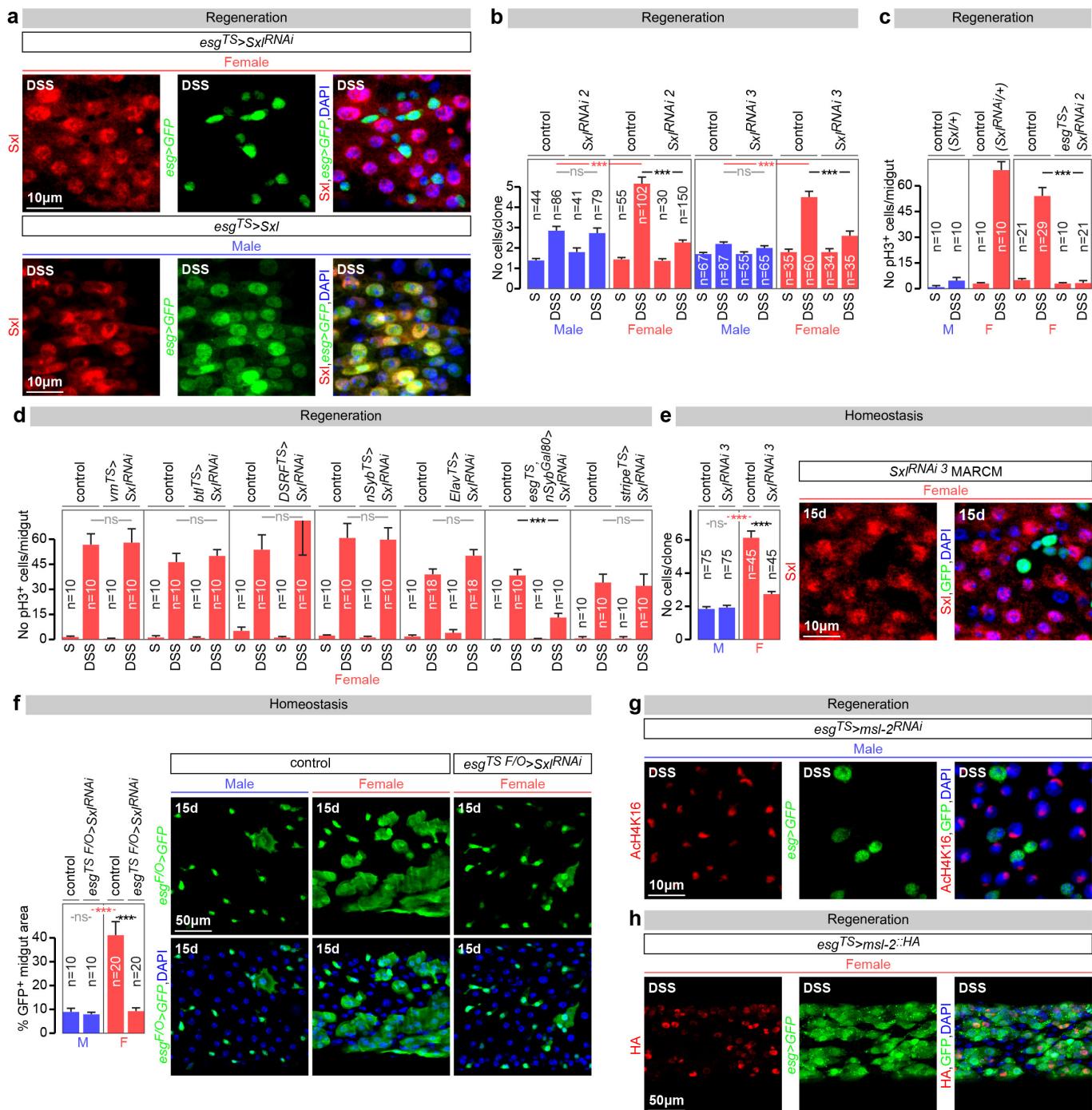
the adult midgut. We note a sexual dimorphism in *dsx* transcript levels. **f**, Venn diagram illustrating the overlap between the genes showing sex-biased expression (overall transcript abundance, light grey, 1,305 genes) and sex-biased alternative splicing (sex-biased isoforms, dark grey, 603 genes) in the adult midgut. Known members of the sex determination pathway are displayed as examples. **g**, Heat maps displaying genes with sexually dimorphic expression clustered by enrichment in specific biological processes, as revealed by Gene Ontology enrichment analysis. Genes with sexually dimorphic expression belonging to the top 4 enriched biological processes are shown. **h, i**, Real-time qRT-PCR data for a subset of genes for which RNA-seq transcriptional profiling experiments revealed sexually dimorphic expression (**h**) or isoforms (**i**). RNA was obtained from midguts from virgin male and female samples (same genotypes as for the RNA-seq experiments). For each gene/isoform, expression abundance was arbitrarily set up at 100% for the sex with the highest expression level, and percentage of that expression is displayed for the other sex. See Methods for details and Supplementary Table 1 for a full list of names and quality scores, and Supplementary Information for full genotypes.



Extended Data Figure 2 | See next page for figure caption.

Extended Data Figure 2 | Cell type-specific expression of sex determinants in the adult intestinal epithelium of virgin flies. **a**, In the adult *Drosophila* midgut, resident stem cells (ISCs) and their postmitotic daughter cells (EBs) maintain the adult intestinal epithelium during normal homeostasis and regenerate it after injury by giving rise to two types of differentiated progeny: ECs and EECs^{80,81}. The posterior midgut area used to visualize and quantify phenotypes is boxed. The following *Gal4* drivers were used to label and/or genetically manipulate these four different cell types present in the adult intestinal epithelium: *mex1* for ECs, *pros^{V1}* for EECs, *esg* for both ISCs and EBs, and *Supressor of Hairless (Su(H))* for EBs alone. **b**, The canonical sex determination pathway in somatic cells of *Drosophila melanogaster*, consisting of a cascade of sex-specific alternative splicing events culminating in the production of sex-specific transcription factors encoded by Dsx and Fru^M. In females, Sex lethal (Sxl) is activated and regulates the splicing of *transformer (tra)* pre-mRNA, resulting in the production of Tra^F. Tra^F regulates the female-specific splicing of *dsx* pre-mRNA (*dsx^F*) and *fru* transcript coming from the P1 promoter (*fru^{P1}*, giving rise to *fru^M*). In males, Sxl is not expressed and no functional Tra is produced, resulting in default splicing of *dsx* and *fru* pre-mRNAs, leading to Fru^M and Dsx^M proteins, respectively. The resulting male- and female-specific Dsx and Fru isoforms confer sexual identity to the cells in which they are produced. In addition, in females, Sxl represses dosage compensation by inhibiting Msl-2 expression. The tables summarize the cell-specific expression profiles of the sex determinants in adult midguts of virgin males and females, shown in the panels below. **c**, Sxl protein (in red) is expressed only in female midguts. Co-staining with ISC/EB reporters indicates that Sxl is found *esg*-positive progenitors (ISCs: GFP-positive and LacZ-negative cells, and EBs: GFP-positive and LacZ-positive cells). It is also expressed in female polyploid ECs

(GFP- (in green) and LacZ- (in blue) negative cells). Co-staining with *pros^{V1}* reporter indicates that it is also expressed in EECs. **d**, Msl-2 protein is found in the same cell types only in males (staining is confined to the X chromosome, consistent with the signal observed in non-intestinal tissues⁸²). **e**, A new reporter of *tra* promoter activity (*tra-LacZ*, see Methods for details) is broadly expressed in the epithelium of both male and female midguts, including ISCs and EBs (as revealed by co-staining with *esg-Gal4*-driven GFP) and ECs (GFP-negative cells with large nuclei). Co-staining with *pros^{V1}* reporter indicates that it is also expressed in EECs. **f**, A *dsx-Gal4* reporter (visualized with a GFP reporter that has been false-coloured in red for consistency with the other panels) is active in male and female polyploid ECs (LacZ-negative cells), and is inactive in *esg*-positive progenitors (LacZ-positive cells, in green). Dsx protein (visualized using a Dsx^{DBD}-specific antibody in red) is expressed strongly only in males but not females, indicating that the sexual dimorphism in *dsx* transcript levels found in our RNA-seq analyses (Extended Data Fig. 1e) is further enhanced at the protein level. Co-staining of the same antibody with the EC marker *mex1-Gal4* confirms expression in ECs. Cytoplasmic *dsx-Gal4*-driven expression of an mTomato reporter is apparent in ECs (visualized as large cells by co-staining with the membrane-enriched marker Armadillo), but is absent from EECs (as revealed by the gaps in mTomato expression in cells that are labelled with Pros. **g**, The *fru^{P1}-Gal4* reporter (which labels the only sexually dimorphic *fru* transcript that gives rise to Fru^M protein) is inactive in both male and female midguts, as revealed by lack of GFP signal (false-coloured in red for consistency with other panels). Consistent with the lack of *fru^{P1}-Gal4* expression, a Fru^M-specific antibody (in red) is not expressed in the male midgut. An independent *dsx* reporter (*dsx^{A2}-Gal4*) is expressed in polyploid ECs, consistent with the data displayed in **f**. See Supplementary Information for full genotypes.

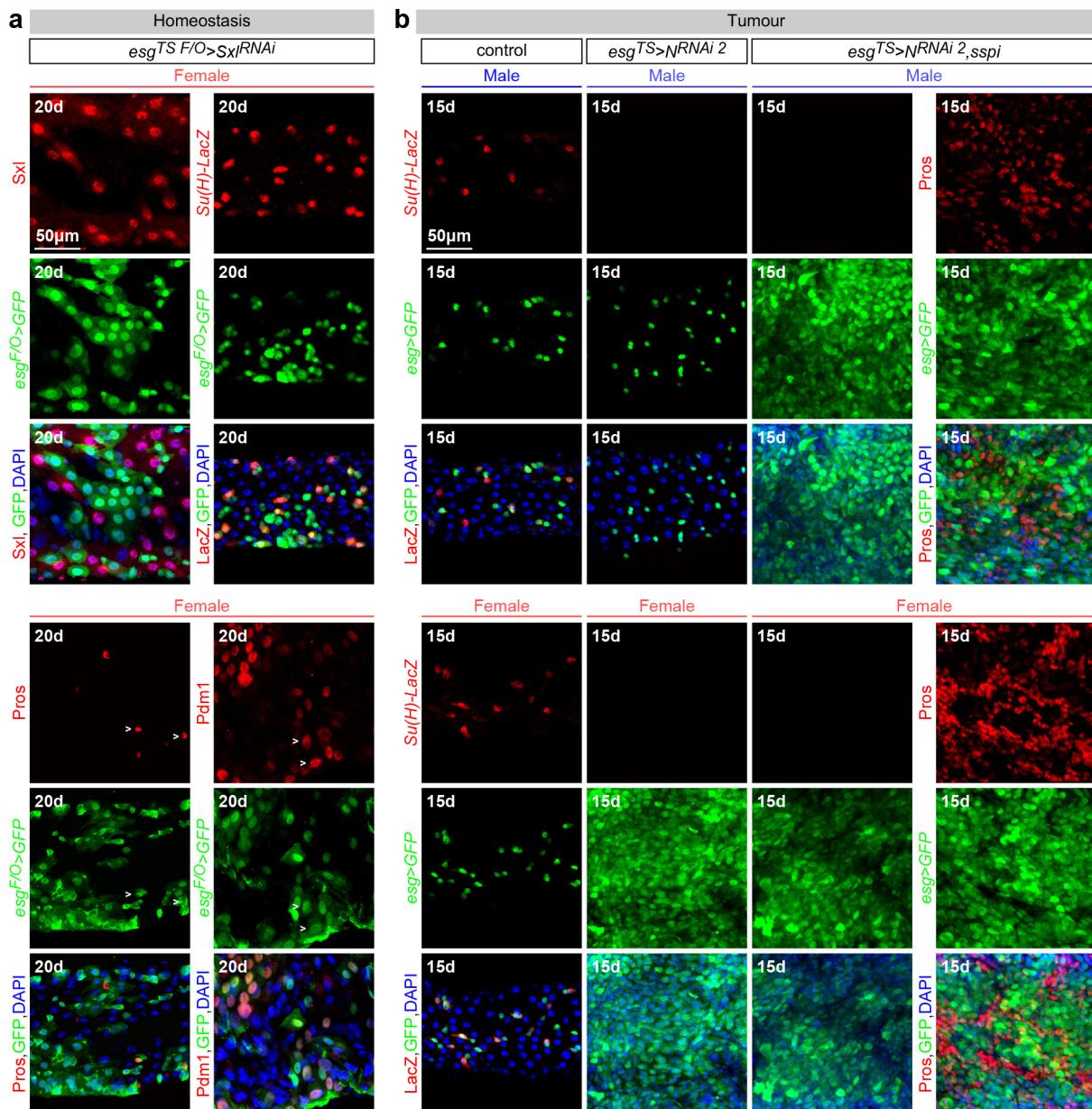


Extended Data Figure 3 | See next page for figure caption.

Extended Data Figure 3 | *Sxl* controls intrinsic sex differences in adult ISC proliferation independently of dosage compensation.

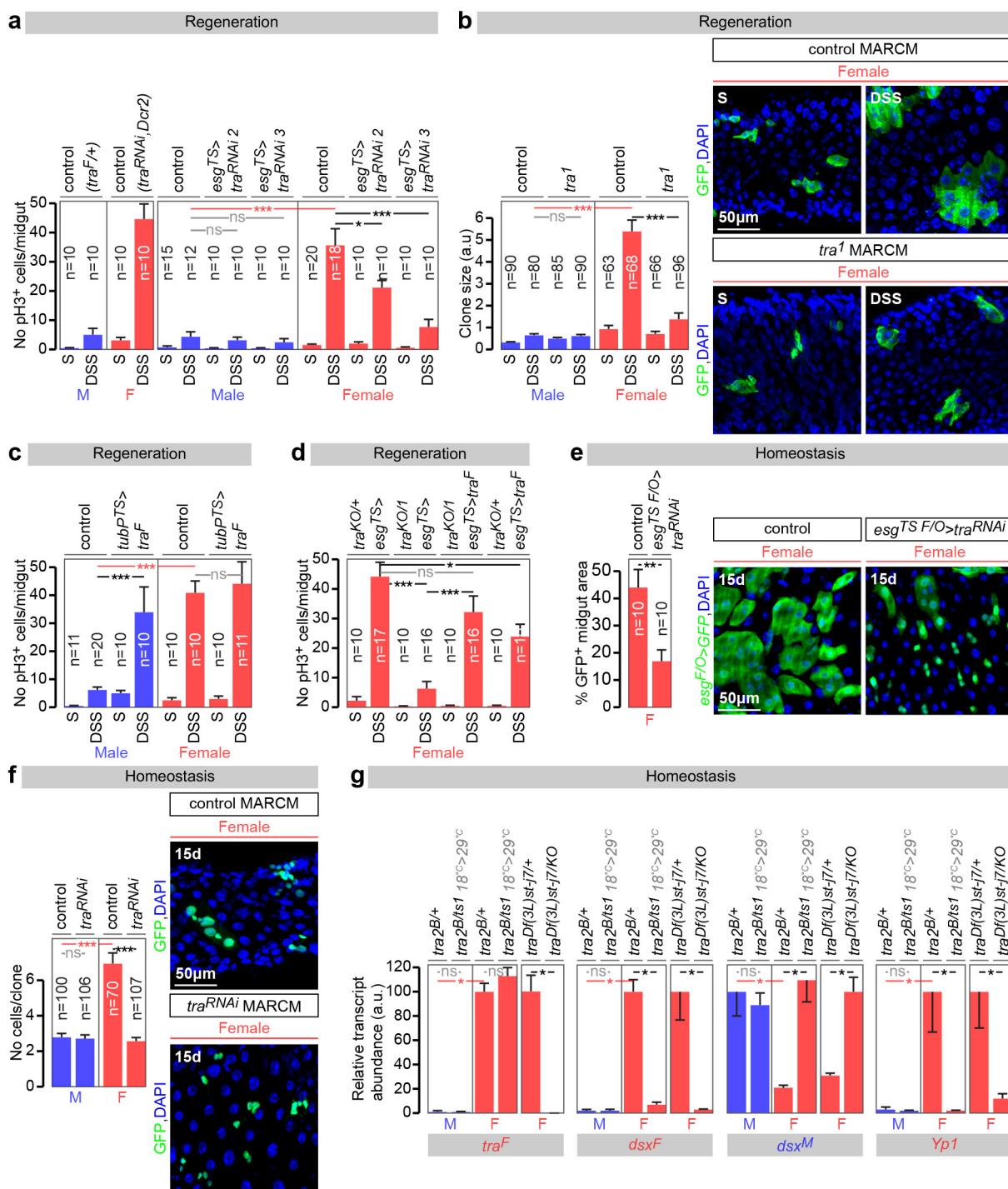
a, Immunocytochemistry using a *Sxl*-specific antibody indicates that adult-restricted downregulation of *Sxl* in intestinal progenitors (ISCs/EBs)—achieved by *Gal80^{TS}*-controlled expression of a *Sxl* RNAi transgene—efficiently downregulates *Sxl* expression in progenitors, but not large polyploid ECs. Conversely, efficient ectopic *Sxl* protein expression is obtained by expression of a *UAS-Sxl* transgene in adult ISCs/EBs of male flies. In all panels *Sxl* antibody is in red; DNA: DAPI, in blue; ISC/EB marker: GFP, in green. **b**, Quantifications of the number of cells inside control MARCM⁸³ clones, or MARCM clones expressing RNAi transgenes directed against *Sxl* following DSS treatment. 5 days after clone induction by heat shock, female clones are larger than male clones only when *Sxl* is present, confirming the cell autonomy of *Sxl* action. **c**, Additional controls for Fig. 1a, and confirmation of phenotypes using an independent RNAi transgene. This second RNAi transgene against *Sxl* (different from the one used in Fig. 1a) reduces the number of pH3-positive cells in DSS-treated female midguts when expressed from *esg^{TS}* in adults ISCs/EBs, confirming an adult progenitor-specific requirement for *Sxl* in promoting damage-induced cell divisions in female flies. **d**, Adult-specific downregulation of *Sxl* in adult visceral muscle (using the *vm* driver), trachea (*btl-Gal4* and *DSRF-Gal4*), neurons (*nSyb-Gal4*, *Elav-Gal4*), or fat body (*stripe-Gal4*) does not reduce DSS-induced ISC proliferation in females. By contrast, *Sxl* downregulation using an ISC/EB driver with suppressed neuronal expression (*esg-Gal4* combined with *nSyb-Gal80*) effectively reduces DSS-induced ISC proliferation in females. Together, these results indicate that *Sxl* does not control sexually dimorphic DSS-induced ISCs proliferation from non-ISC cells. **e**, MARCM clones expressing a third RNAi transgene

against *Sxl* (distinct from those used in Fig. 1 and above) are smaller than control clones in females, whereas their size is comparable to that of wild-type or *Sxl-RNAi* clones in males. This confirms that, during normal homeostasis, female ISCs divide more often than male ISCs because of the cell-autonomous action of *Sxl*. The graph shows quantifications of the number of cells within each clone 15 days after clone induction by heat shock, and the confocal images show representative clones (labelled in green with GFP) for each genotype. **f**, Clonal analyses of homeostatic proliferation using the inducible *esg* flip-out system (which labels progenitors and their progeny generated within a defined temporal window⁸⁴) in midguts of control males, control females and females in which *Sxl* downregulation has been confined to adult progenitors. 15 days after induction, the size (assessed as the percentage of GFP-positive area) of control female clones was significantly larger than that of male clones, but both became comparable upon adult-specific *Sxl* downregulation using *Sxl* RNAi transgenes. The graph shows area quantifications for each sex/genotype, and the confocal images show representative clones for each genotype. **g**, Immunohistochemical detection of histone H4 lysine 16 (H4Lys16) acetylation (in red) indicates that adult-specific downregulation of *msl-2* in male intestinal progenitors (ISCs/EBs marker: GFP, in green) results in loss of H4Lys16 acetylation of the X chromosome. **h**, Efficient *Msl-2* misexpression in adult female intestinal progenitors (ISCs/EBs marker: GFP, in green) is confirmed by immunocytochemistry using an HA-specific antibody (in red). *n* denotes the number of guts (**c**, **d**, **f**) or clones (**b**, **e**) that were analysed for each genotype. Results were combined from at least two independent experiments. See Supplementary Information for full genotypes.



Extended Data Figure 4 | Sexually dimorphic proliferation does not result from sex differences in differentiation. **a**, Markers for all four intestinal cell types are still apparent following adult-specific downregulation of *Sxl* in the intestinal progenitors of females—achieved by *Gal80^{TS}*-controlled expression of a *Sxl* RNAi transgene. Indeed, expression of *esg-Ga4* (ISCs/EBs), *Su(H)-LacZ* (EBs), *Pdm1* (ECs) and *Pros* (EECs) can be readily detected, suggesting that *Sxl* downregulation in females (which results in reduced ISC proliferation) does not have a major effect on differentiation. *Sxl* staining confirms efficient downregulation in ISCs/EBs, but not neighbouring cells. **b**, The same markers reveal that the differentiation defect resulting from *Notch* downregulation, previously reported in females, is also apparent in males (note loss of *Su(H)-LacZ* following *Notch* downregulation in both males and females),

suggesting that sex differences in differentiation do not contribute to the sex differences in susceptibility to *Notch*-induced tumours. Co-expression of a mitogen (secreted *Spitz*, *sSpz*) abrogates the sex differences in tumour susceptibility by efficiently triggering hyperplasia also in males, as revealed by an expanded progenitor (GFP-positive) area in both males and females. The identity of these tumours in males is also comparable to that previously show for *Notch* tumours in females (consisting of high Pro- positive EEC-like cells and low Pro- positive neoplastic ISC-like cells⁸⁵. This further suggests that the sex differences in *Notch*-induced tumour susceptibility do not arise from sexually dimorphic differentiation effects, but result from sex differences in ISC proliferation. See Supplementary Information for full genotypes.

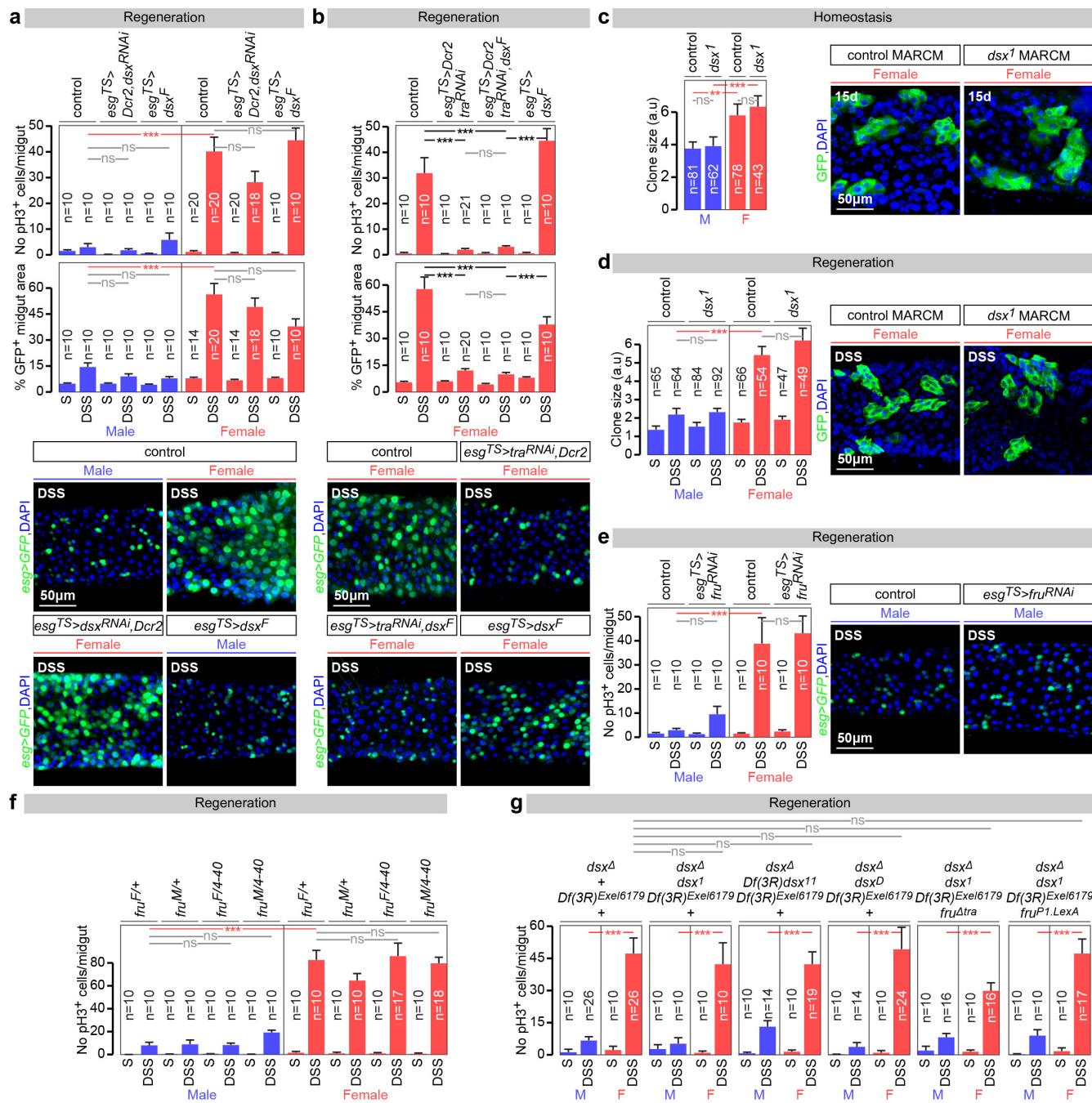


Extended Data Figure 5 | See next page for figure caption.

Extended Data Figure 5 | *tra*, but not *tra2*, controls intrinsic sex differences in adult ISC proliferation.

a, Additional controls for Fig. 2a, and confirmation of phenotypes using independent RNAi transgenes. Two additional RNAi transgenes against *tra* reduce the number of pH3-positive cells in DSS-treated female, but not male, midguts when expressed from *esg*^{TS} in adult ISCs/EBs, confirming an adult progenitor-specific requirement for *tra* in promoting damage-induced cell divisions in female flies. **b**, *tra*¹ MARCM mutant clones are smaller than control clones in females, whereas their size is comparable to that of wild-type or *tra*¹ clones in males. This confirms that female ISCs divide more often than male ISCs because of the cell-autonomous action of *tra*. The graph shows quantifications of clone size (in arbitrary units of GFP fluorescence, as described in Methods) 15 days after clone induction by heat shock, and the confocal images show representative clones (labelled in green with GFP) for each genotype. **c**, Ubiquitous, adult-restricted *tra*^F expression from *tub-Gal80*^{TS} in males increases the number of pH3-positive cells following DSS treatment to levels comparable to those of female flies. **d**, Re-introduction of this *tra*^F transgene specifically in adult ISCs/EBs rescues the reduced, male-like intestinal proliferation (as assessed by the number of pH3-positive cells) of *tra* null mutant females entirely lacking the *tra* gene from all their tissues (*tra*^{KO}/*tra*¹) to levels comparable to those of control females. Expression of this transgene in control heterozygous female flies (*tra*^{KO}/+ *esg*^{TS}>*tra*^F) does not significantly increase their proliferation (in fact, it reduces it slightly relative to *tra*^{KO}/+ *esg*^{TS}> controls, likely as a consequence of its overexpression). **e**, Clonal analyses

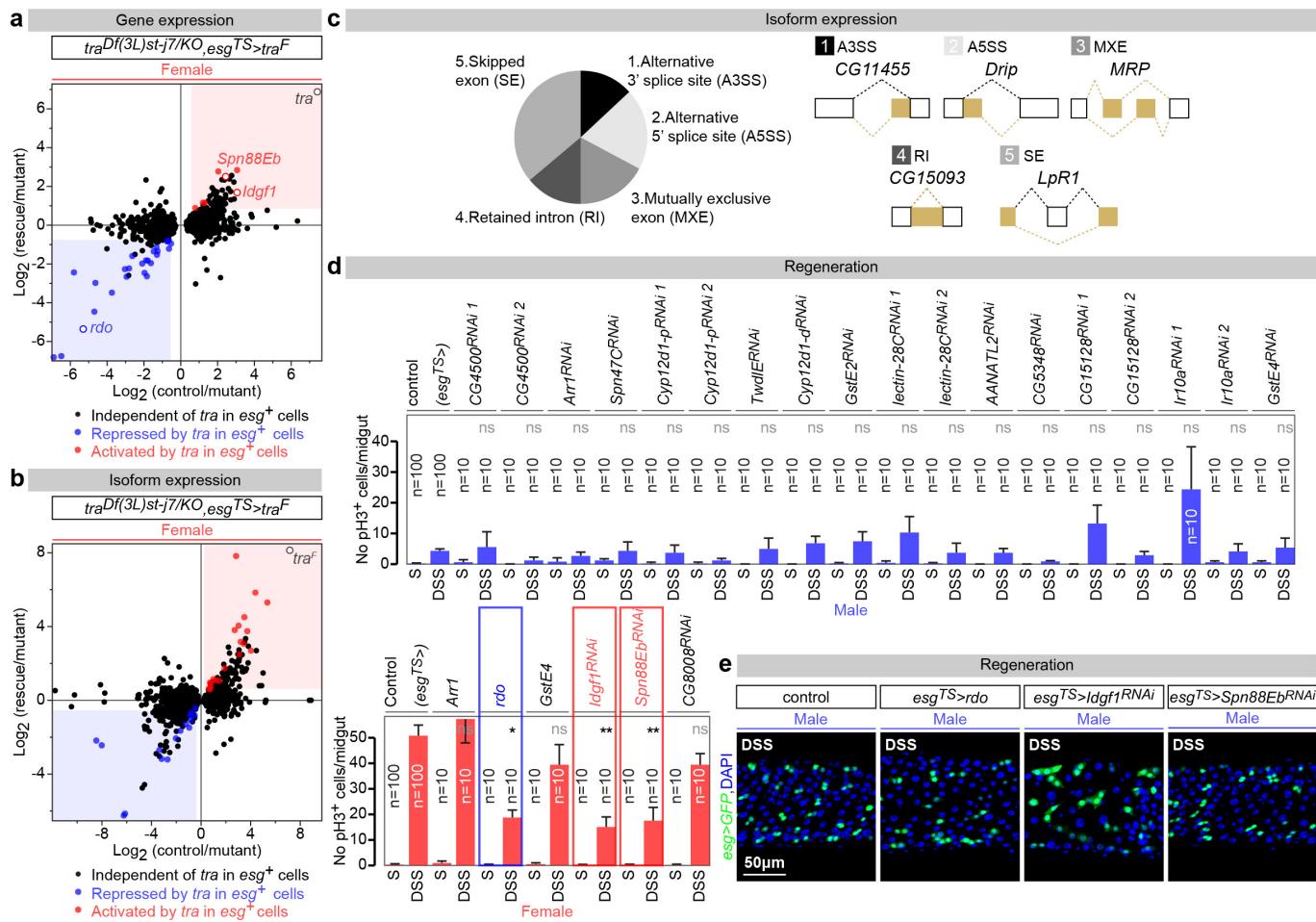
of homeostatic proliferation using the inducible *esg* flip-out system (which labels progenitors and their progeny generated within a defined temporal window⁸⁴) in midguts of control females and females in which *tra* downregulation has been confined to adult progenitors. 15 days after induction, the size (assessed as the percentage of GFP-positive area) of control clones is significantly larger than that of *tra*-RNAi clones. The graph shows area quantifications for each genotype, and the confocal images show representative clones for each genotype. **f**, Consistent with the *tra* mutant clonal analysis in Fig. 2c, quantifications of clone size (number of cells per clone) reveal that MARCM clones in which *tra* has been downregulated are smaller than control clones only in females. Their size is comparable to that of wild-type or *tra*-downregulated mutant clones in males. The confocal images show representative clones (labelled in green with GFP) for each genotype in females. **g**, qRT-PCR quantifications of relative abundance of *tra*^F, *dsx*^F, *dsx*^M and *Yp1* transcripts in adult-specific *tra2* mutants (*tra2*^{B/tst}) grown at permissive temperatures, then switched to the restrictive temperature 4 days after eclosion and transcriptionally profiled following 10 additional days at the restrictive temperature and controls (*tra2*^{B/+}). In *tra2* mutant females, *dsx*^F is lost, *dsx*^M is upregulated to levels comparable to those of control males and *Yp1* (a *Dsx*^F target) is lost (to levels also comparable to those of males). *tra* mutants (*tra*^{Df(3L)st-j7/KO}) were also used as a positive control. *n* denotes the number of guts (**a**, **c**, **d**, **e**) or clones (**b**, **f**) that were analysed for each genotype. Results were combined from at least two independent experiments. See Supplementary Information for full genotypes.



Extended Data Figure 6 | *dsx*- and *fru*^M-independent control of sexually dimorphic proliferation in adult intestinal stem cells.

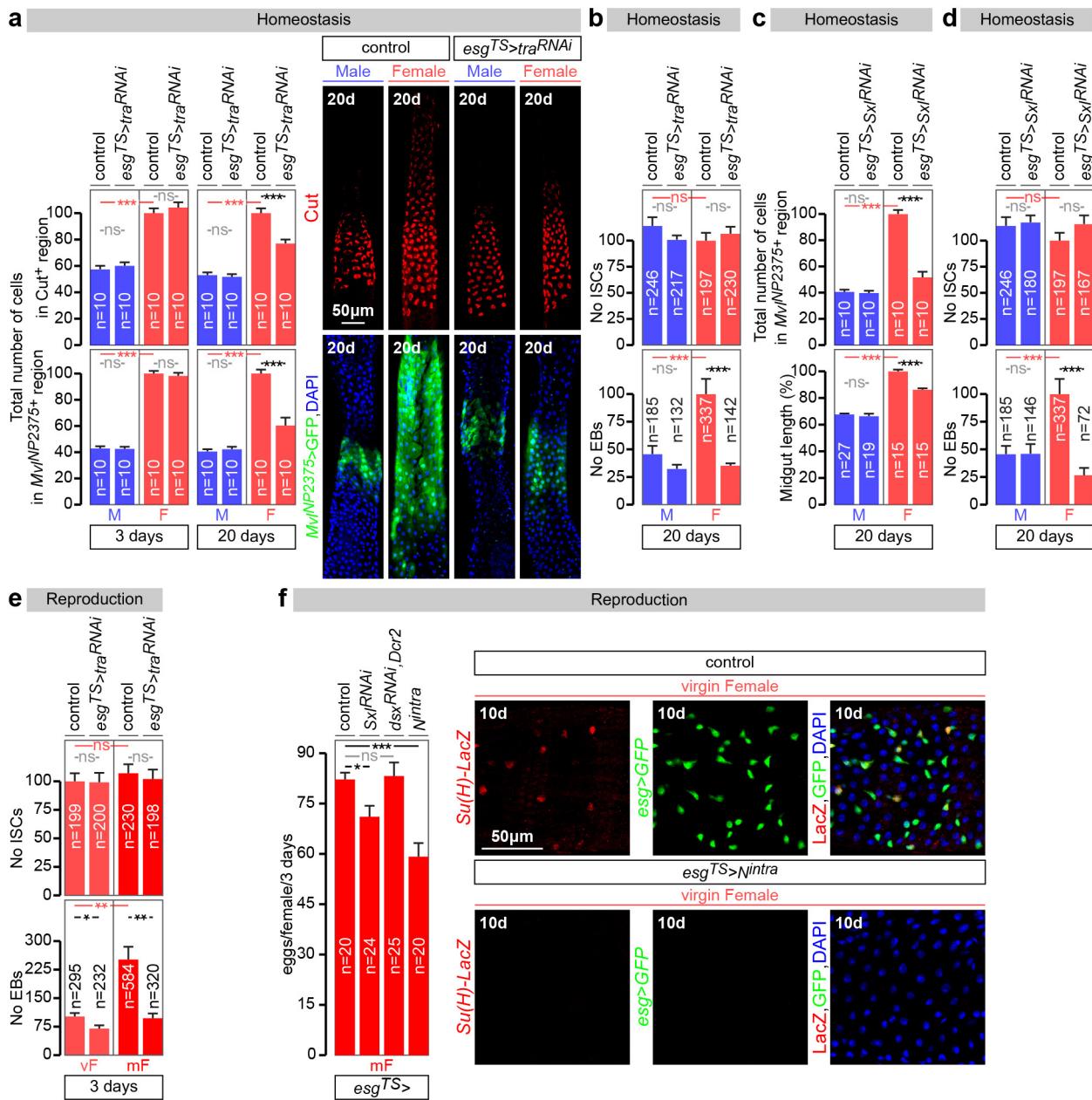
a, Adult-restricted downregulation of *dsx* (achieved by co-expression of a *dsx*-RNAi transgene and *Dicer-2* (*Dcr-2*) in ISC/EBs) has no effect on the compensatory ISC proliferation observed upon DSS treatment in either males or females. *dsx*^F expression in the same conditions does not increase ISC proliferation in either males or females. **b**, *dsx*^F expression does not rescue the reduced proliferation resulting from *tra* downregulation in males. Representative images for each genotype are shown in both **a** and **b** (DNA: DAPI, in blue; ISC/EB marker: GFP, in green). **c**, The size of *dsx* null mutant (*dsx*¹) MARCM clones (quantified in arbitrary units of GFP fluorescence as described in Methods) is comparable to that of controls in both sexes 15 days after clone induction by heat shock. Confocal images show representative clones (labelled in green with GFP) for each genotype. **d**, Quantifications of clone size in control and *dsx* null mutant (*dsx*¹) MARCM clones in the midguts of DSS-treated males and females. 5 days after clone induction by heat shock, there are no significant differences in clone size (quantified in arbitrary units of GFP fluorescence as described in Methods) between control and mutants clones in either males or

females. Confocal images show representative clones (labelled in green with GFP) for each genotype. **e**, An RNAi transgene against *fru* does not reduce the number of pH3-positive cells in DSS-treated midguts when expressed from *esg*^{TS} in the adult ISC/EBs of either males or females. Confocal images show that number of intestinal progenitors (*esg*-positive cells in green) is also unaffected by this manipulation. **f**, Quantifications of the number of pH3-positive cells upon DSS treatment indicates that the sexual dimorphism in ISC proliferation is unaffected in females with forced *fru*^M expression (*fru*^M/*fru*⁴⁻⁴⁰) or in males with forced *fru*^F expression (*fru*^F/*fru*⁴⁻⁴⁰). **g**, ISC proliferation is unaffected in the midguts of DSS-treated males and females entirely lacking *dsx* (*dsx*^Δ/*dsx*¹), or producing only *Dsx*^F (*dsx*^Δ/*dsx*¹¹) or *Dsx*^M (*dsx*^Δ/*dsx*^D). ISC proliferation is also unaffected in *dsx*, *fru*^M double null mutant males and females (*dsx*^Δ, *Df(3R)Exel6179*/*dsx*¹, *fru*^{P1.LexA}), and in *dsx* null mutants in which *fru*^M is ectopically produced in females (*dsx*^Δ, *Df(3R)Exel6179*/*dsx*¹, *fru*^{Δtra}). *n* denotes the number of guts (**a**, **b**, **e**, **f**, **g**) or clones (**c**, **d**) that were analysed for each genotype. Results were combined from at least two independent experiments. See Supplementary Information for full genotypes.



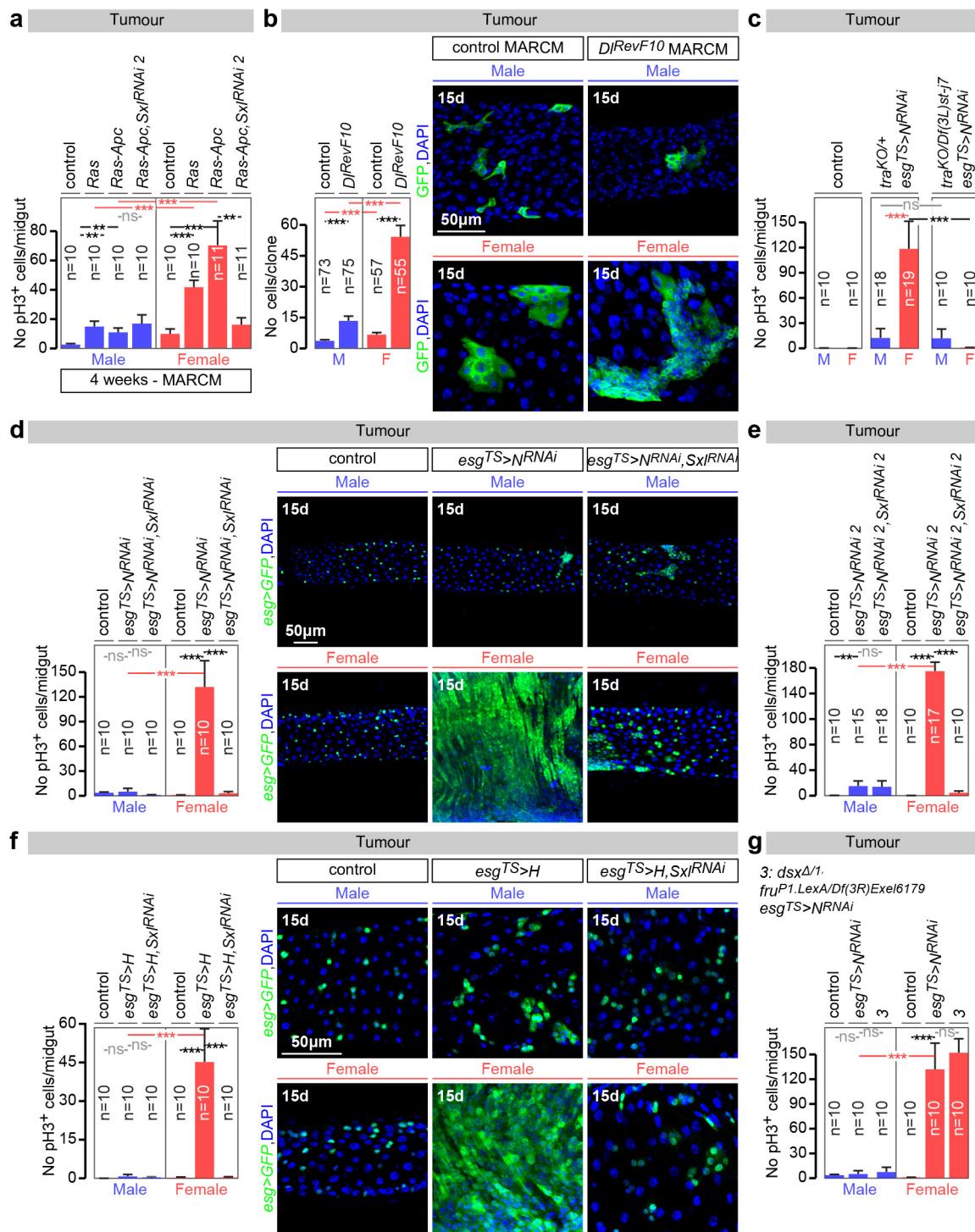
Extended Data Figure 7 | *tra* targets in adult ISCs. **a**, Scatter plot of all 1,346 genes with *tra*-dependent expression in the adult fly midgut. For each gene, control female/*tra* null mutant female (*Df(3L)st-j7/tra*^{KO}) fold differences in transcript abundance (x axis, log₂ scale) are plotted against *tra* mutant female with feminized ISCs (adult-restricted rescue of *tra*^F in ISCs/EBs)/*tra* mutant female fold differences (y axis, log₂ scale). Genes with *tra*-sensitive expression and significantly repressed by *tra*^F in ISC/EBs (*P* < 0.05 cutoff) are therefore found in the left-bottom quadrant and are displayed in blue, whereas those significantly activated by *tra*^F are found in the top-right quadrant and are displayed in red. Genes with *tra*-dependent transcription, independent of the action of *tra*^F in intestinal precursors are displayed in black. **b**, Comparable analysis of *tra*-dependent alternative splicing. **c**, *tra* expression in adult ISCs affects splicing of 38 transcripts by at least 5 different mechanisms. The outcome of each of the alternative splicing mechanisms is shown in yellow for a representative gene. **d**, Adult-restricted downregulation (RNAi lines) or misexpression (UAS lines) of *tra* targets in adult ISCs/EBs by means of *esg-Gal4*, *tubGal80*^{TS}.

Genes normally repressed in female progenitors in a *tra*^F-dependent manner were downregulated in males (top row) and/or misexpressed in females (bottom row). Genes upregulated in female progenitors in a *tra*^F-dependent manner were downregulated in females (bottom row). Adult-restricted downregulation of *Idgf1* and *Spn88Eb* reduces the number of mitoses (pH3-positive cells) in DSS-treated females. Conversely, *rdo* misexpression inhibits DSS-induced ISC proliferation in females. Adult-restricted downregulation of other *tra*^F targets in the same conditions does not affect ISC proliferation in either males or females. **e**, Male controls for Fig. 3c. In contrast to their effects on females, adult ISC/EB-restricted misexpression of *rdo* or downregulation of *Idgf1* and *Spn88Eb* does not reduce the percentage of midgut area covered by *esg*-positive cells in DSS-treated males (DNA: DAPI, in blue; ISC/EB marker: GFP, in green). *n* denotes the number of guts (d) that were analysed for each genotype. Results were combined from at least two independent experiments. See Supplementary Information for full genotypes.



Extended Data Figure 8 | Effects of the sexual identity of adult ISCs on midgut size and reproductive plasticity. **a**, The number of cells in the R3a-b and R4a midgut regions, defined by expression of Cut and *Mv1NP2375* respectively (as described in ref. 86), is higher in females, and can be significantly reduced in females to numbers comparable to those found in males after 20 (but not 3) days of adult-specific downregulation of *tra* in intestinal progenitors (achieved by *esg*^{TS}-driven *tra* downregulation initiated after the phase of midgut post-eclosion growth, see Methods for details). No effects are apparent following downregulation in males. Representative images of these midgut regions (labelled in red with Cut or in green with *Mv1NP2375*, *esg*^{TS}-driven GFP) are shown to the right for each genotype. **b**, Adult ISC/EB-specific *tra* downregulation does not affect the number of ISCs (*esg*-positive, *Su(H)*-negative cells) in either males or females, but strongly reduces EB (*esg*-positive, *Su(H)*-positive cells) production in females. **c**, Quantifications as in **a** and **b** for midguts with adult-specific downregulation of *Sxl* in intestinal progenitors. **c**, Reduced number of cells in the R4a midgut region (top graph) and total midgut length (bottom graph) in female flies following 20 days of adult-specific and cell-autonomous masculinization of their intestinal progenitors (achieved by down regulation of *Sxl* over 20 days with *esg*-Gal4). The same manipulation has no discernible effects in males. **d**, The same genetic manipulation does not affect the number of ISCs (*esg*-positive,

Su(H)-negative cells) in either males or females, but strongly reduces EB (*esg*-positive, *Su(H)*-positive cells) production in females. **e**, The number of EBs (*esg*- and *Su(H)*-positive cells, bottom graph), but not ISCs (*esg*-positive only cells, top graph) is higher in control female flies 3 days after mating. Adult ISC/EB-specific *tra* downregulation abrogates the postmating increase in EBs in females without affecting EB number in males, or ISC number in males or females. **f**, Adult-specific *Sxl* downregulation in intestinal progenitors leads to a small, but significant, reduction in egg production. An unrelated manipulation that also reduces ISC proliferation by inducing differentiation of ISCs (*esg*^{TS}>*Notch*^{intra}, images to the right of the graph and ref. 87) also results in reduced egg production, whereas downregulation of *Dsx* (which does not control sex differences in progenitor proliferation) has no such effect. It should, however, be noted that *esg*-Gal4 is expressed in a subset of cells in the ovary niche²¹. Hence, the possibility that these cells contribute to the observed phenotype cannot entirely be ruled out. Images to the right show loss of intestinal progenitor cell makers *esg*-Gal4 and *Su(H)*-LacZ following expression of *Notch*^{intra} in adult intestinal progenitors, indicative of loss of progenitor identity. *n* denotes the number of guts (**a**, **c**), ISCs/EBs (**b**, **d**, **e**) or female flies (**f**) that were analysed for each genotype. Results were combined from at least two independent experiments. See Supplementary Information for full genotypes.



Extended Data Figure 9 | See next page for figure caption.

Extended Data Figure 9 | Effects of the sexual identity of adult ISC on the susceptibility to genetically induced tumours. **a**, The number of mitoses in *Apc-ras* mutant clones is larger than that of control clones in both sexes, but it is higher and dependent on *Sxl* in females. **b**, The size of *Delta* (*Dl*, the *Notch* ligand) null mutant (*Dl*^{RevF10}) MARCM clones is larger than that of control clones in both sexes, but female mutant clones are larger than male mutant clones. The graph shows quantifications of the number of cells within each clone 15 days after clone induction by heat shock, and the confocal images show representative clones (labelled in green with GFP) for each genotype. **c**, In *tra* ‘female’ mutant flies (*tra*^{KO} and *tra*^{KO}/*Df(3L)*^{st-j7}) reduced *Notch* signalling in intestinal progenitors fails to induce the hyperplasia (quantified by the number of pH3 cells) normally observed in control females. **d**, Following 15 days of adult-specific downregulation of *Notch* (*N*) in intestinal progenitors, hyperplasia (quantified by the number of pH3-positive cells and also shown in representative images) is observed in female, but not male midguts. Adult-specific and cell-autonomous reversal of ISC/EB female identity—achieved by *esg*^{TS}-driven downregulation of *Sxl*—fully prevents the hyperplasia induced by *Notch* downregulation in females, but has no effects on males. Confocal images show intestinal progenitor coverage of representative midgut portions for each genotype (DNA: DAPI, in blue;

ISC/EB marker: GFP, in green). **e**, pH3 quantifications show a comparable effect for independent RNAi transgenes against *Sxl* and *Notch*. **f**, Adult-specific downregulation of *Notch* (*N*) signalling by ectopic expression of the downstream *Notch* signalling antagonist *Hairless* (*H*)³⁸ leads to hyperplasia (quantified by the number of pH3-positive cells and also shown in representative images) in female, but not male midguts. Adult-specific and cell-autonomous reversal of ISC/EB female identity—achieved by *esg*^{TS}-driven downregulation of *Sxl*—fully prevents the hyperplasia induced by *Hairless* overexpression in females, but has no effects on males. Confocal images show intestinal progenitor coverage of representative midgut portions for each genotype (DNA: DAPI, in blue; ISC/EB marker: GFP, in green). **g**, The number of pH3-positive cells 15 days after *Notch* downregulation in adult intestinal progenitors of double null mutant flies lacking *dsx* and *fru*^M (*dsx*^Δ,*Df(3R)*^{Lexel6179}/*dsx*^l,*fru*^{P1,LexA}) is comparable to that of controls in both males and female flies. Like control flies, it is significantly higher in female flies. Virgin flies were used in all these experiments. *n* denotes the number of guts (**a**, **c**, **d**, **e**, **f**, **g**) or clones (**b**) that were analysed for each genotype. Results were combined from at least two independent experiments. See Supplementary Information for full genotypes.

Effector T-cell trafficking between the leptomeninges and the cerebrospinal fluid

Christian Schläger^{1*}, Henrike Körner^{1*}, Martin Krueger³, Stefano Vidoli⁴, Michael Haberl¹, Dorothee Mielke⁵, Elke Brylla³, Thomas Issekutz⁶, Carlos Cabañas⁷, Peter J. Nelson⁸, Tjalf Ziemssen⁹, Veit Rohde⁵, Ingo Bechmann³, Dmitri Lodygin¹, Francesca Odoardi^{1*} & Alexander Flügel^{1,2*}

In multiple sclerosis, brain-reactive T cells invade the central nervous system (CNS) and induce a self-destructive inflammatory process. T-cell infiltrates are not only found within the parenchyma and the meninges, but also in the cerebrospinal fluid (CSF) that bathes the entire CNS tissue^{1,2}. How the T cells reach the CSF, their functionality, and whether they traffic between the CSF and other CNS compartments remains hypothetical^{3–6}. Here we show that effector T cells enter the CSF from the leptomeninges during Lewis rat experimental autoimmune encephalomyelitis (EAE), a model of multiple sclerosis. While moving through the three-dimensional leptomeningeal network of collagen fibres in a random Brownian walk, T cells were flushed from the surface by the flow of the CSF. The detached cells displayed significantly lower activation levels compared to T cells from the leptomeninges and CNS parenchyma. However, they did not represent a specialized non-pathogenic cellular sub-fraction, as their gene expression profile strongly resembled that of tissue-derived T cells and they fully retained their encephalitogenic potential. T-cell detachment from the leptomeninges was counteracted by integrins VLA-4 and LFA-1 binding to their respective ligands produced by resident macrophages. Chemokine signalling via CCR5/CXCR3 and antigenic stimulation of T cells in contact with the leptomeningeal macrophages enforced their adhesiveness. T cells floating in the CSF were able to reattach to the leptomeninges through steps reminiscent of vascular adhesion in CNS blood vessels, and invade the parenchyma. The molecular/cellular conditions for T-cell reattachment were the same as the requirements for detachment from the leptomeningeal milieu. Our data indicate that the leptomeninges represent a checkpoint at which activated T cells are licensed to enter the CNS parenchyma and non-activated T cells are preferentially released into the CSF, from where they can reach areas of antigen availability and tissue damage.

T cells specific for myelin basic protein (MBP) retrovirally transduced to express fluorescent proteins (T_{MBP} cells) were tracked on their way into the CNS tissues after intravenous (i.v.) transfer to healthy recipient rats. The T_{MBP} cells arrived at the CNS on the level of the leptomeninges of the spinal cord before they entered the parenchyma with the onset of clinical disease⁷. Concomitantly to their arrival in the leptomeningeal milieu, T_{MBP} cells accumulated in the CSF (Extended Data Fig. 1a, b).

It has been proposed that immune cells, for example, T helper 17 (T_H17) cells and M2 macrophages, enter the CSF from either the leptomeninges or the choroid plexus before spreading to the CNS tissues^{3–6,8,9}. We could not detect T_{MBP} cells in the choroid vessels or stroma before or during their accumulation in the CSF when we used

intravital two-photon laser scanning microscopy (TPLSM) to image the choroid plexus of the fourth ventricle. Very few cells appeared in the choroid plexus at late phases of CNS infiltration, when T_{MBP} cells had already maximally accumulated in the leptomeninges and CSF (Fig. 1a). Serial fluorescence microscopy of all four choroid plexus, immuno-electron microscopy and cytofluorometric quantifications revealed similar results (Extended Data Fig. 1c–f). To exclude the possibility of transport of T_{MBP} cells from the choroid plexus to the leptomeninges of the spinal cord, we injected a nontoxic polymerizing agent (Matrigel) into the cisterna magna, that is, between the ventricular egression points and subarachnoidal space of the spinal cord⁸. The cisternal block neither prevented the invasion of T_{MBP} cells into the subjacent spinal cord nor influenced the development of clinical disease (Extended Data Fig. 2a, b). Similarly, interrupting the CSF flow from the choroid plexus to the spinal cord did not hinder the accumulation of T_{MBP} cells in the lumbar/thoracic spinal cord (Extended Data Fig. 2c–e).

T_{MBP} cells crawling within the leptomeningeal vessels transgressed the vascular walls (Extended Data Fig. 3a) and moved within the leptomeningeal environment (Supplementary Video 1). This milieu is highly specialized: the pial vessels that run along the surface of the spinal cord parenchyma are suspended in a dense three-dimensional (3D) extracellular matrix (ECM) network consisting mainly of collagen fibrils⁷ (Extended Data Fig. 3b). Just as the bloodstream flows over the vascular endothelia, the CSF within the subarachnoidal space flows over the pial ECM network. Interestingly, when following the movement of single migrating T_{MBP} cells, we regularly observed them at the interface of the ECM network and fluidic compartment becoming detached and being washed into the CSF (Fig. 1b, Extended Data Fig. 3c, Supplementary Videos 2 and 3). This detachment was verified by imaging *in situ* labelled T cells expressing the photo-convertible fluorescent protein Dendra2 (Extended Data Fig. 3d, e, Supplementary Video 4). Photo-converted T_{MBP}-Dendra2 cells in the leptomeninges steadily decreased over time and were replaced by non-photo-converted T_{MBP}-Dendra2 cells, indicating rapid turnover of T cells in this CNS compartment (Extended Data Fig. 3e).

RNA sequencing (RNA-seq) revealed that T_{MBP} cells from the blood, CSF, leptomeninges and spinal cord parenchyma displayed broad conformity of their gene expression profiles, with very similar expression levels of master transcription factors, chemokine receptors, adhesion molecules, cytokine receptors, cell motility and T-cell receptor genes (Fig. 2a, Extended Data Fig. 4a, b, Supplementary Table 1). The exceptions were genes of the T-cell activation program: T_{MBP} cells from the leptomeninges and CNS parenchyma displayed strongly upregulated genes of cytokine–cytokine-receptor interaction,

¹Institute of Neuroimmunology, Institute for Multiple Sclerosis Research, University Medical Centre Göttingen, 37073 Göttingen, Germany. ²Max-Planck-Institute for Experimental Medicine, 37075 Göttingen, Germany. ³Institute of Anatomy, University of Leipzig, 04103 Leipzig, Germany. ⁴Department of Structural and Geotechnical Engineering, University of Rome La Sapienza, 00185 Rome, Italy. ⁵Department of Neurosurgery, University Medical Centre Göttingen, 37075 Göttingen, Germany. ⁶Division of Immunology, Department of Pediatrics Dalhousie University, Halifax B3H 4R2, Canada. ⁷Departamento de Biología Celular e Inmunología, Centro de Biología Molecular Severo Ochoa, 28049 Madrid, Spain. ⁸Medical Clinic and Polyclinic IV, Ludwig-Maximilians-University of Munich, 80336 Munich, Germany. ⁹Department of Neurology, University Hospital, 01307 Dresden, Germany.

*These authors contributed equally to this work.

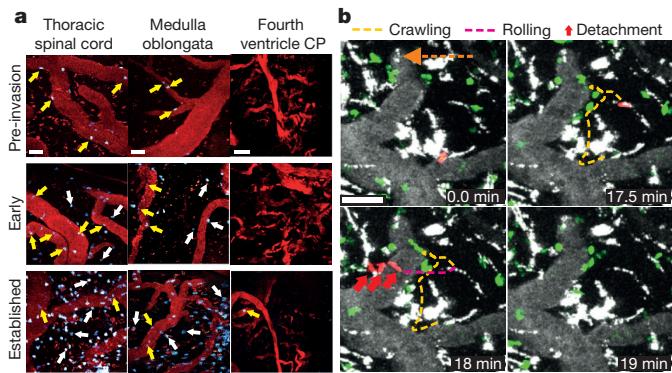


Figure 1 | T_{MBP} cells enter the CSF from the leptomeninges. **a**, Intravital imaging of T_{MBP} cell entry into the leptomeninges and choroid plexus (CP) of the fourth ventricle. TPLSM recordings were performed during the pre-invasion, early and established phases of leptomeningeal T-cell infiltration (see Methods for the definition of phases) sequentially in thoracic spinal cord, medulla and choroid plexus of fourth ventricle in the same animal. Representative recordings of 16 independent experiments. Blue, T_{MBP} -Lifeact-Turquoise2 Cells; red, vessel lumen, meningeal macrophages; yellow and white arrows, intravascular and extravasated cells, respectively. **b**, *In vivo* visualization of T_{MBP} cell detachment from the leptomeninges into the CSF. Intravital images of leptomeninges during the established phase of leptomeningeal T-cell infiltration showing a T_{MBP} -GFP cell (false red colour) detaching from the pial surface. Yellow lines, crawling steps of a T_{MBP} -GFP cell; magenta line, migratory path of a T cell rolling on the leptomeninges and being dragged away into the CSF; grey (false colour), vessel lumen, meningeal macrophages; red arrows, T_{MBP} -GFP cell detachment; orange arrow, CSF flow direction. Scale bars, 50 μ m.

JAK–STAT signalling and T-cell receptor signalling pathways, whereas CSF-derived T_{MBP} cells resembled blood-derived T cells that lacked virtually any signs of activation (Fig. 2b, Extended Data Fig. 4b–e, Supplementary Tables 2 and 3). TPLSM recording of the nuclear translocation of a fluorescently labelled nuclear factor of activated T cells (NFAT-YFP) sensor¹⁰ confirmed these findings: in contrast to the leptomeninges and the CNS parenchyma where 26% and 35%, respectively, of the T_{MBP} -NFAT/YFP cells showed nuclear translocations of the sensor, virtually no T_{MBP} -NFAT/YFP cells in the CSF displayed nuclear NFAT translocations (< 2%; Fig. 2c).

The low activation profile of CSF-derived T_{MBP} cells could not be explained by inhibitory effects of the CSF on T-cell activation (Extended Data Fig. 4f). Most notably, CSF-derived T_{MBP} cells were fully responsive to antigenic stimulation and induced ‘classic’ EAE (Fig. 2d), indicating that T cells in the CSF—despite their reduced activation levels—were fully functional.

We next performed i.v. transfer of myelin oligodendrocyte glycoprotein (MOG)-reactive T cells (T_{MOG} cells), which enter the CNS tissues but reach only low reactivation levels within the CNS and induce very mild clinical disease¹¹ (Extended Data Fig. 5a). T_{MOG} cells invaded the CNS parenchyma to a lesser extent than T_{MBP} cells but remained in higher percentages in the CSF and leptomeninges (Fig. 2e and Extended Data Fig. 5b). This shifted distribution pattern became even more pronounced when brain non-reactive ovalbumin-specific T cells (T_{OVA} cells) were transferred into rats. They entered in the CNS milieu similarly to T_{MBP} or T_{MOG} cells, though in much lower numbers (Fig. 2e), but mostly remained in the CSF/leptomeninges compartments (Fig. 2e and Extended Data Fig. 5b) and did not evoke any detectable CNS inflammation (Extended Data Fig. 5c).

The situation markedly changed in inflamed CNS tissue. When T_{OVA} cells were co-transferred with pathogenic T_{MBP} cells that induced a strong upregulation of chemokines, integrin ligands and cytokines within the leptomeninges and parenchyma, high numbers of T_{OVA} cells, irrespective of their low activation levels, invaded the CNS and migrated from the CSF/leptomeninges compartments deep into the

CNS parenchyma (Fig. 2e and Extended Data Fig. 5b–d). There were still more T_{OVA} cells in the CSF than T_{MBP} cells, indicating that even during CNS inflammation the lack of T-cell reactivation affects the adhesiveness of T_{OVA} cells within the leptomeningeal milieu (Extended Data Fig. 5e). Taken together, these data indicate that in addition to the T-cell reactivation state, inflamed CNS tissue supports T-cell attachment to the leptomeninges and migration into the CNS parenchyma.

We next investigated the cellular partners and molecular cues that guide T cells within the leptomeningeal milieu. Enmeshed in the pial ECM are numerous macrophages that scan their environment by protruding and retracting cellular processes (Extended Data Fig. 6a). T_{MBP} cells crawling through the leptomeningeal milieu were in contact with these cells the majority of the time (> 70%) (Extended Data Fig. 6a, b, Supplementary Video 5). In contrast to T cells in lymphatic tissue¹², the T_{MBP} cells in the leptomeninges moved in a Brownian random walk (Fig. 3a, b) not following measurable microgradients (Extended Data Fig. 6c). CD8⁺ T cells searching for rare infectious foci during toxoplasma infection of the CNS displayed a Lévy walk behaviour that is characterized by a combination of random walk and large jumps¹³. Apparently when the interacting partners within the CNS tissue are abundant, as is the case with leptomeningeal macrophages, Brownian motion is sufficiently efficient¹⁴. Meningeal macrophages express major histocompatibility complex (MHC) class II molecules on their surface and therefore can act as antigen-presenting cells^{10,15,16}. In addition, they express fibronectin and ICAM-1 (Extended Data Fig. 6d), that is, ligands of VLA-4 ($\alpha 4\beta 1$) and LFA-1 ($\alpha L\beta 2$) integrins expressed by T_{MBP} cells (Extended Data Fig. 7a). *In situ* staining with an antibody directed against the activated conformation of VLA-4 revealed that 10% of the T_{MBP} cells within the leptomeninges, but none in the CSF, carried activated integrin on their membrane (Extended Data Fig. 7b). Intrathecal (i.t.) injection of neutralizing anti-VLA-4 and/or anti-LFA-1 monoclonal antibodies shortened the contact time of T_{MBP} cells with leptomeningeal phagocytes, whereas the contact frequency increased (Fig. 3c). Furthermore, in the presence of the CSF flow, the velocity of the T cells increased (Fig. 3c, Extended Data Fig. 7c, d), but their random walk behaviour remained unaffected (Extended Data Fig. 7e). Interestingly, we noted a large leptomeningeal detachment of T_{MBP} cells that then rolled along the pial surface or floated in the CSF (Fig. 3d, e, Extended Data Fig. 7f, Supplementary Video 6). Accordingly, photo-converted T_{MBP} -Dendra2 cells vanished more rapidly from the leptomeninges after VLA-4/LFA-1 interference (Extended Data Fig. 7g).

Chemokines are known to activate the integrin-mediated binding of immune cells¹⁷. T_{MBP} cells upregulate chemokine receptors before entering the CNS, in particular CCR5, CXCR3 and CXCR4 (ref. 18; Extended Data Fig. 8a). The corresponding chemokines (CCL5, CXCL9–11 and CXCL12) were present in the CNS tissues and specifically expressed by the resident macrophages (Extended Data Fig. 8b–d). After i.t. application of pertussis toxin (PTX; which acts as a global chemokine inhibitor), high numbers of T_{MBP} cells became detached from the pial surface and were dragged into the CSF (Extended Data Fig. 8e, Supplementary Video 7). Their random walk behaviour and straightness of movement were unchanged but their contact time with the resident macrophages decreased (Extended Data Fig. 8f, g). In contrast to integrin blockade, the T-cell velocity remained unchanged (Extended Data Fig. 8f). Similar observations were made after i.t. injection of Met-RANTES and anti-CXCR3 monoclonal antibodies, which specifically interfere with CCR5 and CXCR3, respectively. In contrast, the CXCR4 blocker AMD3100 had no effect (Fig. 3f, Extended Data Fig. 8f, h).

T_{OVA} cells also migrated in a Brownian locomotion pattern and came into contact with the local macrophages, but these contacts were less frequent and shorter than those of T_{MBP} cells and their velocity was higher (Extended Data Fig. 9a–c). Interfering with integrins or chemokines also caused a strong detachment of T_{OVA} cells from the leptomeninges (Extended Data Fig. 9d, e). Notably, neither integrin

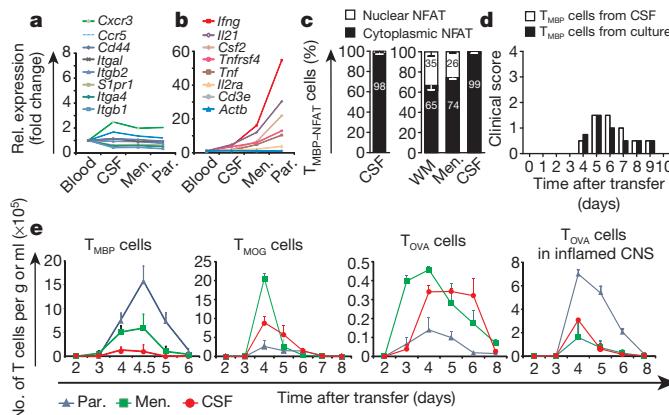


Figure 2 | Characterization of T_{MBP} cells in the CSF. **a**, Similarities in transcriptome profiles between T_{MBP} cells in blood, CSF, leptomeninges and CNS parenchyma. RNA-seq was performed for T_{MBP}-GFP cells sorted from the indicated compartments 3 days after transfer, and relative (rel.) mRNA expression of selected adhesion molecules and chemokine receptor genes is shown. **b**, Genes related to T-cell activation are upregulated in T_{MBP} cells from meninges and CNS parenchyma compared to CSF. Relative mRNA expression (RNA-seq analyses) of selected genes that are highly regulated in T_{MBP}-GFP cells from the indicated CNS compartments. Expression levels in the T cells isolated from blood was set to 1 (a, b). House-keeping genes, *Actb* and *Cd3e*. **c**, *In vivo* analyses of T_{MBP}-NFAT cells confirm that their activation levels within the leptomeninges exceed those in the CSF. Percentages of cells with either cytoplasmic (not activated) or nuclear (activated) NFAT sensor during the established phase of leptomeningeal T-cell infiltration were determined by TPLSM (218 cells from 5 independent experiments, left plot) or by histology of SC cryosections (white matter (WM): 352 cells, meninges: 1,371 cells) and of cytocentrifuge preparations (CSF: 326 cells), right plot. Data are mean \pm s.d. **d**, T_{MBP} cells in the CSF retain their pathogenic potential. 1.7 \times 10⁶ *in vitro* reactivated T_{MBP}-GFP cells isolated from CSF of animals 3 days after transfer or an equivalent number of *in vitro* T cell blasts were injected i.v. into naive animals. Representative clinical score of 2 independent experiments ($n = 6$). The moderate severity of clinical disease is due to the low amount of injected T cells. **e**, T-cell reactivation and inflammatory state of the CNS tissue determine T-cell trafficking from leptomeninges into the CSF or CNS parenchyma. Flow cytometry quantification of antigen-specific T cells in the CNS compartments at the indicated time points after transfer. The first three plots show T_{MBP}-GFP cells, T_{MOG}-GFP cells and T_{OVA}-GFP cells with high, low or no reactivation potential, respectively. Right, T_{OVA}-GFP cell distribution in an inflamed CNS tissue (co-transfer with non-labelled T_{MBP} cells). Data are mean \pm s.e.m. of at least 6 animals per group per time point for each antigen specificity from at least 2 independent experiments.

nor chemokine interference influenced the basal locomotion characteristics or the Brownian motility pattern of T_{OVA} or T_{MBP} cells (Extended Data Figs 8f, g, and 9f–h). These findings indicate that effector T cells in the 3D network of the leptomeninges—similarly to dendritic cells in a 3D environment¹⁹—are integrin-independent and follow an intrinsic locomotion program independent of chemokinetic or chemoattractive stimuli. In contrast, the interactions with local antigen-presenting cells are regulated by chemokine- and integrin-mediated adhesive forces that prevent effector T cells from being released into the CSF from the flow-exposed 2D milieu of the leptomeninges.

T cells were previously proposed to migrate from the CNS tissues via the CSF along lymphatic vessels of the dura mater^{20,21}, but it is unclear whether T cells in the CSF also return to the leptomeninges and the CNS parenchyma. After i.t. injection, T_{MBP} cells were indeed spread within the leptomeninges and the adjacent parenchyma (Extended Data Fig. 10a). This infiltration pattern resembled that of early EAE lesions but clearly differed from that of solutes²², which were proposed to be transported from the CSF into the CNS parenchyma via peristaltic forces along periarterial spaces. This ‘glymphatic’ transport

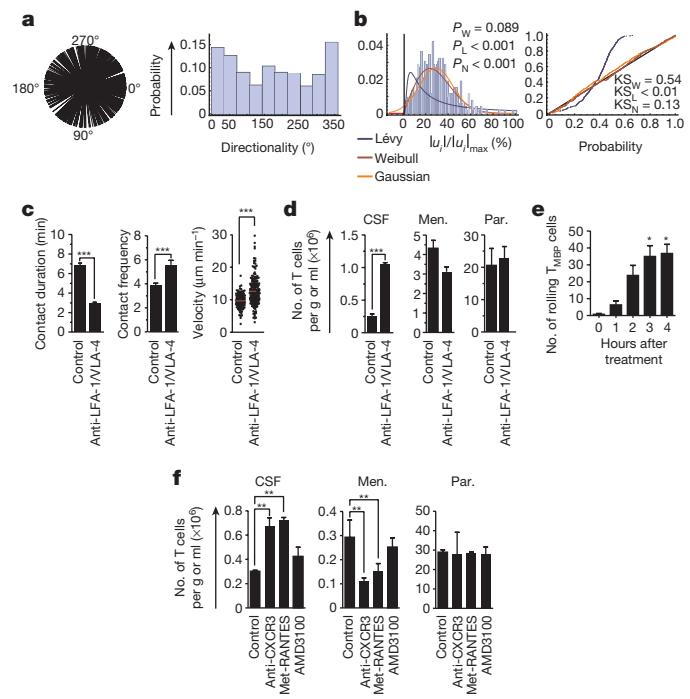


Figure 3 | Integrins and chemokines mediate T_{MBP} cell adhesion to leptomeningeal structures. **a, b**, TPLSM recordings show the motility behaviour of T_{MBP}-GFP cells in the leptomeningeal milieu. Time interval, 32 s. **a**, T_{MBP} cells do not follow a preferential direction. Directions (left) and associated probability for 0–360° angles (right). **b**, T_{MBP} cells move in a Brownian random walk. Left, probability distribution of T-cell displacements in a linear scale with overlapping fitting of the displacement norm ($|\mathbf{u}_i|$) for Lévy (blue), Weibull (brown) and normal (Gaussian, orange) distributions. Respective P values are indicated. Only the Weibull distribution (equivalent to Brownian movement) has statistical significance. Right, probability plots of the indicated distributions against the displacement norm data with the associated results of Kolmogorov–Smirnov (KS) test. Analysis of at least 9 TPLSM recordings. **c**, Integrin blockade interferes with T cell and antigen-presenting cell contact and accelerates T-cell migration. Intravital 30 min TPLSM recordings. Mean contact duration (529 contacts per 208 cells), contact frequency (208 cells) between T_{MBP}-GFP cells and leptomeningeal phagocytes and velocity (320 cells) before (control) and 4 h after i.t. injection of LFA-1/VLA-4-blocking monoclonal antibodies. **d, e**, Integrin blockade induces a release of T_{MBP} cells from the meninges into the CSF and increases T-cell rolling. **d**, Flow cytometry quantification of T_{MBP}-GFP cells in the different CNS compartments 4 h after i.t. treatment with PBS (control) or blocking antibodies. Data are mean \pm s.e.m. of 3 independent experiments (**c, d**) including three animals per group (**d**) (two-tailed Mann–Whitney U -test). **e**, Number of T_{MBP}-GFP cells rolling on the pial surface during 10 min TPLSM recordings before (0 h) and at the indicated time points after integrin blockade. **f**, Interference with chemokine signalling induces a release of T_{MBP} cells from the leptomeninges into the CSF. Flow cytometry quantification of T_{MBP}-GFP cells in the different CNS compartments 4 h after i.t. treatment with PBS (control), anti-CXCR3 monoclonal antibody, Met-RANTES or AMD3100. Data are mean \pm s.e.m. of representative data from two (**e**) or three (**f**) independent experiments including three animals per group (Kruskal–Wallis ANOVA followed by Dunn’s multiple comparison test). Analyses performed during the established phase of leptomeningeal T_{MBP} cell infiltration (**a–f**). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ (**c–f**).

mechanism does not seem to apply to cellular components²³. We detected that T-cell transport in the CSF was mainly driven by the animal’s respiration rather than the cardiac cycle²⁴ (Extended Data Fig. 10b, Supplementary Video 8). After injection of T_{MBP} cells into the subarachnoidal space of the cisterna magna or the lumbar spinal cord, the majority of cells accumulated at the levels of cell injections, that is, at the medulla/cervical and lumbar/thoracic spinal cord, respectively

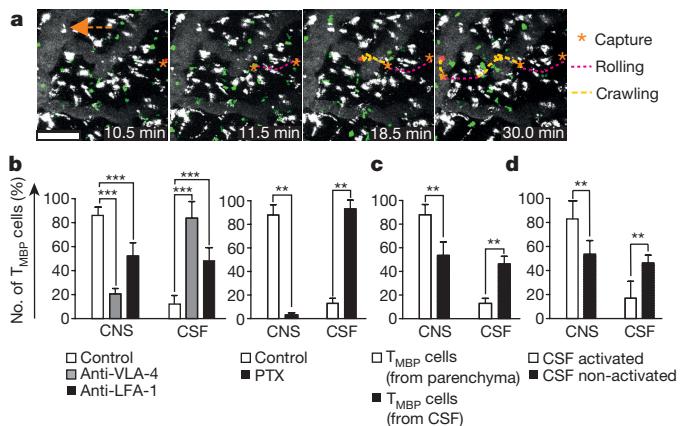


Figure 4 | Reattachment of effector T cells from the CSF to the leptomeninges and entry into the CNS parenchyma. **a,** *In vivo* visualization of TMBP cell reattachment from the CSF to the leptomeninges. TPLSM recordings (30 min) during the established phase of leptomeningeal TMBP cell infiltration. Example of a TMBP-GFP cell (false red colour) reattaching from the CSF to the pial surface with subsequent rolling (magenta dotted lines), capture (orange asterisks) and crawling steps (yellow dotted lines). Green, TMBP-GFP cells; grey (false colour), vessel lumen, meningeal macrophages; orange arrow, direction of CSF flow. Time interval, 32 s. Scale bar, 50 μ m. **b–d,** Conditions that regulate TMBP cell reattachment and entry into the CNS tissue. **b,** Blocking of integrin or $\text{G}\alpha_i$ signalling strongly reduces the capacity of TMBP cells to migrate from the CSF into the CNS parenchyma. TMBP-GFP cells isolated from the CNS parenchyma were pre-treated with anti-VLA-4 or anti-LFA-1 monoclonal antibodies or with isotype control IgG (control) (left), or PTX or PBS (control) (right). The cells were then injected i.t. into naive animals. Entry of the transferred TMBP-GFP cells into the spinal cord (including meninges and parenchyma) and the CSF was quantified by flow cytometry 42 h later, that is, when the re-transferred T cells had maximally infiltrated the spinal cord of the recipient animals. Mean \pm s.d. of representative results of 3 independent experiments ($n = 18$, Kruskal–Wallis ANOVA followed by Dunn's multiple comparison test (left) or two-tailed Mann–Whitney U -test, right). **c, d,** Activation fosters trafficking of TMBP cells from the CSF into the CNS tissue. **c,** TMBP-GFP cells isolated from CSF or CNS parenchyma were injected i.t. into naive animals ($n = 6$). **d,** CSF-derived TMBP-GFP cells with or without previous antigenic stimulation *in vitro* were transferred i.t. Cell quantifications as in **b**. Data are mean \pm s.d. of three (**c**) and two (**d**) independent experiments for each condition ($n = 8$, two-tailed Mann–Whitney U -test). ** $P < 0.01$, *** $P < 0.001$ (**b–d**).

(Extended Data Fig. 10c). However, a considerable number of cells also reached remote areas of the CNS, indicating that the CSF can serve as transport medium for the cells. These distributions clearly differ from that of regular EAE (Extended Data Fig. 10d), in agreement with our observation that during EAE effector T cells enter the meningeal milieu via local vessels.

TMBP cells initially floating within the CSF rolled along the pial surface and then stopped; thereafter, they continued their locomotion by crawling or rolling along the leptomeningeal structures (Fig. 4a, Supplementary Videos 4, 9 and 10), reminiscent of vascular adhesion steps^{7,25}. The molecular requirements for the attachment of the T cells were also similar and included integrins and chemokines (Fig. 4b). However, the molecular range in leptomeningeal adhesion included LFA-1 in addition to VLA-4 (Fig. 4b), only the latter being a requirement for T-cell adhesion to the CNS endothelium⁷. Furthermore, integrin blockage completely abolished T-cell rolling along the endothelium¹⁸, but not along the pial surface (Fig. 3e, Supplementary Video 6), indicating that during the leptomeningeal reattachment integrins control fixed adhesion of the T cells rather than their first initial binding steps. Another notable difference to vascular adhesion was the role of T-cell activation. CNS endothelial cells do not act as antigen-presenting cells for the incoming TMBP cells¹⁰. However, antigen-driven T-cell activation significantly contributed to the T-cell reattachment in

the leptomeninges: non-activated CSF-derived TMBP-GFP cells entered the CNS tissue after i.t. injection less efficiently than their activated counterparts from the CNS parenchyma (Fig. 4c). When the CSF-derived T cells were activated before transfer, their invasive capacity significantly increased (Fig. 4d). Consequently, i.t.-transferred activated TMBP cells invaded the CNS parenchyma and induced clinical EAE, whereas the resting cells were non-pathogenic (Extended Data Fig. 10e). Pretreatment of the activated T cells with anti-integrin monoclonal antibodies significantly reduced the severity of clinical EAE (Extended Data Fig. 10f). Tova cells attached less efficiently than TMBP cells. However, when the Tova cells were activated, after antigenic stimulation before i.t. transfer or by evoking antigen-specific reactivation within the CNS by supplying OVA i.t.²⁶, they invaded the CNS more efficiently (Extended Data Fig. 10g).

The reattachment to the leptomeninges was not only dependent on the activation state of the T cells but also on the inflammatory state of the CNS tissue, as after i.t. transfer, TMBP cells preferentially invaded inflamed CNS tissue compared to that of control rats (Extended Data Fig. 10h). The enhanced invasion could be substantially reverted by pretreating the cells with integrin/chemokine blockers before i.t. transfer (Extended Data Fig. 10i). Both observations indicate that the molecular/cellular requirements for reattachment of T cells from the CSF to the leptomeningeal milieu were the same as for their adhesion.

It is important to note that our findings were not an artefact of transfer EAE. In animals with active EAE induced by reactivated memory TMBP cells²⁷, T cells exhibited a very similar trafficking behaviour and function in the CSF (Extended Data Fig. 10j–l).

Our data support the view that the leptomeninges represent an important checkpoint for T-cell infiltration of the CNS during autoimmune inflammation or immune surveillance^{7,16,28}. Trafficking between CSF and CNS tissue is well regulated by integrin adhesive forces that are triggered by T-cell activation and/or chemokines^{17,29} (Extended Data Fig. 10m), indicating that the CSF fulfils a dual role: it functions as a depot that prevents potentially dangerous effector T cells from entering the CNS parenchyma if these cells had unsuccessfully scanned the meninges for antigens or damage; and the circulating CSF can be used by T cells for residence or as transport medium, similar to blood circulation, to rapidly travel to damaged areas of the CNS. T cells in the CSF displayed very similar expression profiles to effector T cells that had invaded the CNS tissue. Furthermore, they maintained full antigen responsiveness and pathogenic potential. Therefore, the characterization of readily accessible T cells in the CSF could be of relevance for gaining insights into the properties and function of pathogenic T cells in multiple sclerosis.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 26 October; accepted 17 December 2015.

Published online 10 February 2016.

1. Alvermann, S., Hennig, C., Stuve, O., Wiendl, H. & Stangel, M. Immunophenotyping of cerebrospinal fluid cells in multiple sclerosis: in search of biomarkers. *JAMA Neurol.* **71**, 905–912 (2014).
2. Giunti, D. *et al.* Phenotypic and functional analysis of T cells homing into the CSF of subjects with inflammatory diseases of the CNS. *J. Leukoc. Biol.* **73**, 584–590 (2003).
3. Kivisäkk, P. *et al.* Human cerebrospinal fluid central memory CD4⁺T cells: evidence for trafficking through choroid plexus and meninges via P-selectin. *Proc. Natl. Acad. Sci. USA* **100**, 8389–8394 (2003).
4. Reboldi, A. *et al.* C-C chemokine receptor 6-regulated entry of T_H-17 cells into the CNS through the choroid plexus is required for the initiation of EAE. *Nature Immunol.* **10**, 514–523 (2009).
5. Engelhardt, B. & Ransohoff, R. M. The ins and outs of T-lymphocyte trafficking to the CNS: anatomical sites and molecular mechanisms. *Trends Immunol.* **26**, 485–495 (2005).
6. Engelhardt, B., Wolburg-Buchholz, K. & Wolburg, H. Involvement of the choroid plexus in central nervous system inflammation. *Microsc. Res. Tech.* **52**, 112–129 (2001).
7. Bartholomäus, I. *et al.* Effector T cell interactions with meningeal vascular structures in nascent autoimmune CNS lesions. *Nature* **462**, 94–98 (2009).

8. Shechter, R. *et al.* Recruitment of beneficial M2 macrophages to injured spinal cord is orchestrated by remote brain choroid plexus. *Immunity* **38**, 555–569 (2013).
9. Baruch, K. & Schwartz, M. CNS-specific T cells shape brain function via the choroid plexus. *Brain Behav. Immun.* **34**, 11–16 (2013).
10. Lodygin, D. *et al.* A combination of fluorescent NFAT and H2B sensors uncovers dynamics of T cell activation in real time during CNS autoimmunity. *Nature Med.* **19**, 784–790 (2013).
11. Kawakami, N. *et al.* The activation status of neuroantigen-specific T cells in the target organ determines the clinical outcome of autoimmune encephalomyelitis. *J. Exp. Med.* **199**, 185–197 (2004).
12. Castellino, F. *et al.* Chemokines enhance immunity by guiding naïve CD8⁺ T cells to sites of CD4⁺ T cell–dendritic cell interaction. *Nature* **440**, 890–895 (2006).
13. Harris, T. H. *et al.* Generalized Lévy walks and the role of chemokines in migration of effector CD8⁺ T cells. *Nature* **486**, 545–548 (2012).
14. Humphries, N. E. *et al.* Environmental context explains Lévy and Brownian movement patterns of marine predators. *Nature* **465**, 1066–1069 (2010).
15. Mues, M. *et al.* Real-time *in vivo* analysis of T cell activation in the central nervous system using a genetically encoded calcium indicator. *Nature Med.* **19**, 778–783 (2013).
16. Kivisakk, P. *et al.* Localizing central nervous system immune surveillance: meningeal antigen-presenting cells activate T cells during experimental autoimmune encephalomyelitis. *Ann. Neurol.* **65**, 457–469 (2009).
17. Laudanna, C., Kim, J. Y., Constantin, G. & Butcher, E. Rapid leukocyte integrin activation by chemokines. *Immunol. Rev.* **186**, 37–46 (2002).
18. Odoardi, F. *et al.* T cells become licensed in the lung to enter the central nervous system. *Nature* **488**, 675–679 (2012).
19. Lämmermann, T. *et al.* Rapid leukocyte migration by integrin-independent flowing and squeezing. *Nature* **453**, 51–55 (2008).
20. Louveau, A. *et al.* Structural and functional features of central nervous system lymphatic vessels. *Nature* **523**, 337–341 (2015).
21. Aspelund, A. *et al.* A dural lymphatic vascular system that drains brain interstitial fluid and macromolecules. *J. Exp. Med.* **212**, 991–999 (2015).
22. Iliff, J. J. *et al.* A paravascular pathway facilitates CSF flow through the brain parenchyma and the clearance of interstitial solutes, including amyloid β . *Sci. Transl. Med.* **4**, 147ra111 (2012).
23. Hatterer, E., Touret, M., Belin, M. F., Honnorat, J. & Nataf, S. Cerebrospinal fluid dendritic cells infiltrate the brain parenchyma and target the cervical lymph nodes under neuroinflammatory conditions. *PLoS ONE* **3**, e3321 (2008).
24. Dreha-Kulaczewski, S. *et al.* Inspiration is the major regulator of human CSF flow. *J. Neurosci.* **35**, 2485–2491 (2015).
25. Ley, K., Laudanna, C., Cybulsky, M. I. & Nourshargh, S. Getting to the site of inflammation: the leukocyte adhesion cascade updated. *Nature Rev. Immunol.* **7**, 678–689 (2007).
26. Odoardi, F., Kawakami, N., Klinkert, W. E., Wekerle, H. & Flügel, A. Blood-borne soluble protein antigen intensifies T cell activation in autoimmune CNS lesions and exacerbates clinical disease. *Proc. Natl Acad. Sci. USA* **104**, 18625–18630 (2007).
27. Kawakami, N. *et al.* Autoimmune CD4⁺ T cell memory: lifelong persistence of encephalitogenic T cell clones in healthy immune repertoires. *J. Immunol.* **175**, 69–81 (2005).
28. Lucchinetti, C. F. *et al.* Inflammatory cortical demyelination in early multiple sclerosis. *N. Engl. J. Med.* **365**, 2188–2197 (2011).
29. Dustin, M. L. & Springer, T. A. T-cell receptor cross-linking transiently stimulates adhesiveness through LFA-1. *Nature* **341**, 619–624 (1989).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors thank S. Hamann, A. Stas, N. Meyer, S. Mole, and M. Weig for excellent technical assistance. We thank G. Salinas-Riester for her support in performing the transcriptome analyses, T. Lingner for his help in analysing the transcriptome data and W. Lühder for contributing to the mathematical T-cell locomotion analyses. We are grateful to C. Ludwig for text editing. This work was supported by the Deutsche Forschungsgemeinschaft (TRR-SFB43 project B10, FORR 1336 project B1 and RK-Grant FL 377/3-1), the Bundesministerium für Bildung und Forschung ('UNDERSTAND MS'), the Hertie Foundation (grants 1.01.1/11/004 and 1130072), the Ministry of Science and Culture of Lower Saxony (Niedersachsen-Research Network on Neuroinfectiology, N-RENNT) and the European Commission ERA-NET NEURON (MELTRA-BBB).

Author Contributions C.S. performed most intravital TPLSM imaging studies. H.K. performed most of the CSF analyses, fluorescence microscopy and T-cell-CSF-transfers. M.K., E.B. and I.B. performed and analysed immunofluorescence and electron microscopic analyses. S.V. performed the mathematical cell motility analyses. M.H. performed antibody labelling and contributed to TPLSM imaging. D.M. and V.R. designed the operative strategy, D.M. performed the plexus preparation. C.C. provided HUTS4 antibody and contributed with technical advice. T.I. provided the anti-VLA-4 and anti-CXCR3 antibodies and contributed with technical advice. P.J.N. provided CCR5 blocker and contributed with scientific advice. T.Z. contributed with T-cell characterization in the CSF. D.L. designed and produced genetic retroviral sensors and contributed to the analysis of NSeq data. F.O. performed most *ex vivo* T-cell analyses, i.e. cytofluorometric characterizations, quantitative PCR and NSeq analyses. A.F. together with F.O. designed the study, coordinated the experimental work and wrote the manuscript with inputs from co-authors.

Author Information The data discussed in this publication have been deposited in the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus and are accessible through GEO Series accession number GSE75488. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.F. (fluegel@med.uni-goettingen.de).

METHODS

Animals. Rats on a LEW/Crl background (*Rattus norvegicus*) were bred in the animal facilities of the University Medical Center, Göttingen (Germany) and held under standardized conditions. All animal experiments were approved by the responsible authorities (number: 209.1/211-2531-36/04 and 33.9.42502-04-016/09). Male and female animals between 6–12 weeks old were used in the EAE experiments. No differences were noted between the sexes.

Generation and culturing of T cells. CD4⁺ T cells reactive against myelin basic protein, recombinant myelin oligodendrocyte glycoprotein (amino acids 1–120) or ovalbumin were retrovirally engineered to express eGFP or mCherry (denoted as T_{MBP}-GFP, T_{OVA}-GFP, T_{MOG}-GFP and T_{MBP}-Cherry) as previously reported³⁰. The generation of MSCV-NFAT/YFP-Cherry/H2B transduced T-cell lines (T_{MBP}-NFAT/YFP) is described elsewhere¹⁰. For the generation of T_{MBP}-Lifeact-Turquoise2 or T_{MBP}-Dendra2 cell lines, the fragment coding for the fusion protein Lifeact-mTurquoise2 (Addgene, plasmid 36201) or for the photo-switchable protein Dendra2³¹ were cloned into the MCS of the murine stem cell retrovirus pMSCVpuro (Invitrogen). The generation of primary effector T-cell lines has been previously described³⁰. Extraction of guinea pig MBP and production of recombinant MOG were performed as previously described^{32,33}. Ovalbumin (albumin from chicken egg white grade V) was obtained from Sigma. All T-cell lines were CD4⁺, CD8⁻ and $\alpha\beta$ TCR⁺, they had an effector memory phenotype (L-selectin⁻CD45RC^{low}CD44^{high}) and upon stimulation produced IFN γ and IL-17. Phenotype, cytokine profile, antigen specificity, pathogenicity and absence of mycoplasma contamination were verified in each cell line³⁰.

EAE models. Adoptive transfer EAE was induced by i.v. injection of 5×10^6 MBP- or MOG-reactive T-cell blasts (day 2 after antigen encounter). In some experiments $2-5 \times 10^6$ *ex vitro* or $0.5-2 \times 10^6$ *ex vivo* isolated T_{MBP} cells were transferred into the cisterna magna. To address the role of CNS-non-reactive T cells in a non-inflammatory situation, 5×10^6 T_{OVA}-GFP/mCherry blasts were transferred as above. To investigate the behaviour of T_{OVA} cells in inflamed CNS, 2.5×10^6 T_{OVA}-GFP cells were co-injected with 5×10^6 non-labelled T_{MBP} cells. For *in vivo* recording during the established phase of leptomeningeal T cell infiltration, 2×10^6 T_{MBP}-GFP blasts were co-injected with 3×10^6 non-labelled T_{MBP} cells. This protocol allowed the accurate tracking of individual effector T cells.

Active EAE was induced in 8- to 12-week-old memory animals by subcutaneous antigen immunization. Memory animals were established by intraperitoneal transfer of T_{MBP}-GFP cells into neonatal animals as previously described²⁷. Weight and clinical scores were recorded daily (score 0, no disease; 1, flaccid tail; 2, gait disturbance; 3, complete hind limb paralysis; 4, tetraparesis; 5, death).

No statistical method was used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Cell isolation, flow cytometry and fluorescence-activated cell sorting. Isolation of fluorescently labelled T cells from the tissues has been previously described³⁴. Briefly, mononuclear cells were isolated from EDTA-treated blood by density gradient. Immune cells were obtained from spinal cord meninges and parenchyma by a two-phase Percoll-density gradient. CSF was collected from the cisterna magna using a stereotactic device. The choroid plexus of the fourth and lateral ventricles was excised under a stereomicroscope by carefully pulling it off the wall of the ventricles along the meninges. Nevertheless, it cannot be excluded that adjacent meningeal tissue was excised together with the choroid plexus, thus leading to an overestimation of T_{MBP} cell numbers. Meningeal macrophages were labelled by injection of (3 kDa) Texas-Red-conjugated dextran in the cisterna magna (40 μ g per rat) and isolated by density gradient 24 h later. Labelled cell populations were sorted by a FACS Aria 4L SORP cell sorter (Becton Dickinson). Flow cytometry analysis was performed with a FACSCalibur operated by Cell Quest software (Becton Dickinson). Surface staining was performed by using the following mouse anti-rat monoclonal antibodies: OX-40 antigen (CD134), OX-39 antigen (CD25), IL-2 receptor α chain, CD11b/c (OX-42), $\alpha\beta$ TCR-AP647 (clone R73, BioLegend), TCR V β 8.2 (clone R78, Santa Cruz Biotechnology), TCR V β 8.5 (clone B73), TCR V β 10 (clone G101), TCR V β 16.1 (clone His42), CD4-PECy7 (clone OX-35, BD Biosciences), CD62L-PE (clone OX-85, BioLegend), CD45RC-PerCPCy5.5 (clone OX-22, Abcam). All antibodies were purchased from Serotec unless indicated otherwise. Mouse IgG1 κ (MOPC 31C, Sigma-Aldrich) served as isotype control; APC-labelled anti-mouse antibody (Dianova) was used as secondary antibody. For detection of the activated conformation of integrin β 1, HUTS-4 antibodies³⁵ (0.5 mg kg⁻¹) were injected i.t. After 4 h, *ex vivo* isolated T_{MBP}-GFP cells were stained with the secondary antibody. For intracellular staining, mouse anti-rat IFN γ antibodies (DB-1, Invitrogen) and rat anti-mouse IL-17-PE (BD) were used. Mouse IgG1 κ (MOPC 31 C) and rat IgG1-PE were used as control. Cytotfluorometric quantification of T cells was performed by relating the number of cells to a known absolute amount of fluorescent beads.

Intravital TPLSM. Intravital studies were performed during the following phases of leptomeningeal T_{MBP} cell infiltration: (1) pre-invasion, when T_{MBP} cells first appeared in leptomeningeal blood vessels but had not yet extravasated; (2) early, at the beginning of the leptomeningeal T-cell infiltration, that is, when T_{MBP} cells started to distribute within the leptomeninges but had not yet invaded the CNS parenchyma (the animals were still completely healthy); (3) established, when T_{MBP} cells had entered the leptomeninges and started to invade the CNS parenchyma (the animals began to lose body weight); (4) disease, when T_{MBP} cells invaded the CNS parenchyma (the animals showed symptoms of paralytic disease).

Surgical procedures. Animals were anaesthetized, tracheally intubated, ventilated and stabilized in a custom-made microscope stage; body temperature was regulated by a heated pad (37.5 °C). During imaging, vital parameters were registered as previously described⁷. Thoracic leptomeninges were accessed as previously described⁷ by performing a laminectomy at level Th12/L1. Brainstem area was exposed as described³⁶. For recording the choroid plexus, a cerebellar interhemispheric approach was performed to access the fourth ventricle. The rat was placed in a prone position with the head flexed and fixed in a custom-made stereotactic frame. The skin was incised in the midline (length approximately 2 cm) in order to expose the posterior cranial vault and first cervical vertebra. The muscle was detached using regular small scissors and bipolar forceps. The atlanto-occipital membrane was exposed between the foramen magnum and the first cervical vertebra. Afterwards, the skull was thinned out with a twist drill up to the inner cortical layer. This layer was then removed with a slightly curved forceps. Subsequently, the brainstem and the vermis were exposed and could be visualized through the dural layer. The dura was opened in the midline in a blunt fashion using bipolar forceps or a blunt micro-hooklet. The adjacent brain tissue was retracted to the side and partially reduced in size by bipolar coagulation and resection in a 30–45° angle upwards in the sagittal plane in order to enter the fourth ventricle. Upon ventricle entry, CSF and the basal part of the fourth ventricle plexus could be visualized. After each preparation fluorescently labelled dextran was injected i.v. to confirm that the choroid plexus vessels were patent.

Labelling of phagocytic cells and blood vessels. Meningeal phagocytes were labelled by i.t. injection of 3 kDa Texas-Red-conjugated Dextran 48 h after transfer. Blood vessel lumen was visualized by i.v. infusion of (2000 kDa) TRITC-labelled dextran (200 μ g) before or during TPLSM recordings.

Technical equipment and processing of raw data. Time-lapse TPLSM was performed using a LSM710/Axio Examiner Z1 microscope (Carl-Zeiss Microimaging) combined with a > 2.5 Watt Ti:Sapphire Chameleon Vision II Laser device (Coherent GmbH). The excitation wavelength was tuned to 880 nm or 1010 nm and routed through a $20 \times$ water NA1.0 immersion objective W Plan Apochromat (Carl-Zeiss Microimaging). Typically, areas of $424.27 \times 424.27 \mu\text{m}$ (512 \times 512 pixels) width were scanned and 50–100 μm z-stacks were acquired. The interval time between sequential acquisitions was kept to 32 s. Emitted fluorescence was detected using non-descanned detectors (Carl-Zeiss Microimaging) equipped with 442/46 nm, 483/32 nm, 525/50 nm, 550/49 nm, 607/70 nm and 624/40 nm band-pass filters (Semrock Inc.). Collagen was detected by two-photon generated second-harmonic signals.

Irreversible photo-conversion of T_{MBP}-Dendra2 cells from green to red was achieved by irradiation of meningeal spots with UV light for 20 s. Recordings were performed at 880 nm using 483/32 nm and 550/49 nm band-pass filters. TPLSM recordings were acquired and processed by Zen 2009 Software (Carl-Zeiss Microimaging). The depth colour-coding of individual T cells in the z-plane was generated in Fiji software.

Analysis of T-cell motility. Imaris software 7.1.1 (Bitplane) was used for 3D reconstructions and 4D analysis of acquired raw data. Cells were tracked using the automated Imaris Track module with subsequent manual revision. Motility parameters including T-cell velocity (average of the instantaneous speed for each cell track), crawling duration and meandering index (ratio between total T-cell path length and the sum of the entire single displacements) were calculated as described^{7,37} within a 30 min recording interval. Rolling T cells were defined as cells appearing as single or several round-shaped dots or cells moving solely in direction of the blood flow with $> 50 \mu\text{m min}^{-1}$ instantaneous velocity. Owing to limitations in temporal resolution of the two-photon scanning/fluorescence video microscopy equipment used, fast rolling events might have escaped our analysis⁷. The motility parameters were analysed before and after each individual treatment. The value registered before the treatments did not show any difference between the groups and therefore they were pooled together and indicated in the legends as control. Statistical evaluations were performed with GraphPad 5.0.4 (GraphPad Software). **Analysis of T-cell interactions with meningeal phagocytes/analysis of T-cell activation *in vivo*.** Meningeal phagocytes were labelled as above. We exclusively analysed motile GFP⁺ T cells and spots with similar density of fluorescently labelled phagocytes. Contact durations were determined by manually counting the number of

frames during which individual motile GFP⁺ T cells were in close vicinity (≤ 1 cell diameter distance) to resident phagocytes. As not all T cells were visible during the entire observation period of 30 min, contact frequencies were calculated as follows: the total number of phagocytes contacted by an individual T cell was divided by the total number of T-cell displacements. The obtained value was extrapolated to 30 min. For evaluation of T-cell activation *in vivo*, T cells with nuclear (translocated) or cytosolic (not translocated) NFAT sensor were quantified as previously described¹⁰ by analysing fluorescent overlap between the green and red channel. Merged (yellow), translocated; not merged, not translocated.

Mathematical analysis of chemotactic gradients. Target migration was analysed as follows. At first, the positions of individual motile T cells and meningeal macrophages from individual TPLSM recordings were extracted. For every T_{MBP-GFP} cell entering a radius of 25, 50 or 100 μm from a macrophage all the motile steps were analysed. For each step we computed the direction relative to the examined phagocyte. In particular we computed the cosine between the normalized vectors AB and AC as shown in Extended Data Fig. 6c (the letter A corresponding to the initial position of the T_{MBP-GFP} cell, B to its final position and C to the position of the phagocyte). Hence, the values 1 or -1 correspond to a T_{MBP-GFP} cell moving towards the meningeal phagocyte or in the opposite direction, respectively.

Analysis of T-cell migration. *Analysis of the directionality of T-cell displacement.* This analysis was performed by using standard scripts on three different time scales, the fastest sampling time ($\Delta t = 32$ s), and two slower times, $4\Delta t$ and $8\Delta t$, respectively. The three analysed time scales gave the same results; therefore, exclusively data for the fastest time scale, that is, the larger data sets, are reported.

For analysing T-cell movement in the leptomeninges, 3D displacement components were recorded. The directional analysis was split into two parts. Specifically, we first found the plane fitting the positions of all of the effector cells at all times; we then computed the normal and in-plane components of each displacement vector \mathbf{u}_i ,

$$w_i = \mathbf{u}_i \cdot \mathbf{n}, \quad v_i = \mathbf{u}_i - (\mathbf{u}_i \cdot \mathbf{n})\mathbf{n}$$

The directional migration analysis was performed separately on these two components. Here \mathbf{n} is the unit normal to the best fitting plane.

Particularly for the sets of displacement normal components w_i , we estimated the Wiener processes maximizing the log-likelihood function. The cell displacements normal to the best-fitting plane were well approximated by Wiener processes with very small positive drifts; the typical drift was 5–8% of the average normal displacement norm, that is, $\bar{w} = \langle |w_i| \rangle$. This means that the cells moved in a random way without following any preferential direction in the \mathbf{n} direction.

Analysis of effector T-cell movement type. The distribution of the displacement norm $|\mathbf{u}_i|$ was analysed. The histograms of the probability density functions for the recorded data together with the estimated Weibull, Lévy and normal (Gaussian) distributions are shown on linear scales; for all the three cases this estimation was obtained by a global maximization of the log-likelihood function. For each distribution the associated P values are reported. To compare the three hypotheses, probability plots were generated in which all of the estimated distributions were plotted against the data; the ideal line would be a straight line from (0; 0) to (1; 1). For all the cases investigated, the Weibull distribution was the most significant. This result was also confirmed by Kolmogorov–Smirnov test. Hence, for our data, the tail of the displacement-normalized distribution decayed exponentially (Weibull distribution) and it was not well approximated by a polynomial decay typical of the Lévy distribution. Considering that (1) there was no evidence of preferential directions, and (2) the displacement-norm tail decayed in an exponential way, we concluded that the effector T cells moved in a Brownian random walk.

Interference with integrin signalling. Intravital studies were performed during the early and established phases of leptomeningeal TMBP infiltration. To block integrin-mediated binding, a neutralizing mouse anti-rat monoclonal antibody against VLA-4 (anti-CD49d, clone TA-2)³⁸ and/or against CD11a (integrin α L, anti-LFA-1, clone WT.1, Serotec) was applied in the cisterna magna before or during intravital imaging at a single dose of 1 mg kg⁻¹. After recording, *ex vivo* isolated T cells were tested by flow cytometry for antibody saturation.

The effect of integrins on T cell motility in the leptomeninges in the absence of flow was investigated removing the dura mater and the subarachnoida during intravital imaging. The subarachnoidal structures were left intact. Then, 30 min TPLSM recordings were performed before and after *in situ* application of VLA-4 and LFA-1 monoclonal antibodies (20 μg each per rat).

Interference with chemokine signalling. Intravital studies were performed during the early and established phases of leptomeningeal T_{MBP} infiltration. Pertussis toxin A oligomer (10 μg kg⁻¹) (PTX, an exotoxin that blocks G-protein-coupled G α i signalling and therefore acts as global chemokine inhibitor; List biological laboratories), 1 mg kg⁻¹ Met-RANTES, 1 mg kg⁻¹ hamster anti-rat CXCR3

monoclonal antibody (clone XR3.2), 0.5 mg kg⁻¹ AMD3100 (also known as plerixafor) (Sigma/Genzyme) or PBS were applied in the cisterna magna before or during intravital imaging. These monoclonal antibodies or blocking agents were shown to be effective *in vivo* in EAE models^{39–41}. The dose selected for the *in vivo* experiments was able to block T-cell migration in chemotaxis assays. In some experiments, pertussis toxin B oligomer, Armenian hamster IgG or mouse IgG1 isotype antibody (Abcam) were used as controls. To confirm the efficacy of the treatments after the imaging session, *ex vivo* isolated effector T cells were tested in chemotaxis assays towards the respective chemokines.

Intravital fluorescence microscopy. For fast-acquisition fluorescence video microscopy, a LSM710/Axio Examiner Z1 microscope (Carl-Zeiss Microimaging) was used in combination with an HXP120C illuminator routed through a 20 \times water NA1.0 immersion objective W Plan Apochromat (Carl-Zeiss Microimaging). Acquisition rate was 7–10 frames per minute. Fluorescent signals were detected using a Zeiss AxioCam HSM video camera. Mechanical hyper-ventilation was induced by increasing the ventilation rate in 200 g rats from 81 to 100 breaths per minute (bpm) equivalent to the ventilation rate of 100 g animals. Cardiac vagolytic effect was achieved by subcutaneous administration of the muscarinic receptor antagonist methylscopolamine (0.05 mg kg⁻¹) 30 min before imaging⁴².

Block of T-cell transport from the fourth ventricle to the subarachnoidal space. Two different procedures were performed. CSF flow obstruction was obtained by injecting growth-factor-reduced Matrigel (BD Bioscience; 160 μl for 140–170 g animals) intracisternally as previously described^{8,43}. To assess the success of the Matrigel obstruction, we mixed Evans blue dye into the Matrigel and visually confirmed that the Matrigel not only enclosed the medulla oblongata but also completely filled the entire fourth ventricle (Extended Data Fig. 2a, left panel). Visual inspection of Matrigel localization in and around the fourth ventricle was performed in all animals of each experiment.

A spinal cord window was opened at the cervical level before intravital imaging. Then, the dura was removed and a small hole inserted in the arachnoidea, allowing the CSF to leak out. The CSF efflux induced a collapse of the subjacent thoracic subarachnoidal space and thereby abolished the CSF flow in this area (Extended Data Fig. 2c, d).

Intrathecal injection of T cells. Effector T cells were either cultured *in vitro* and injected on day 2 (activated cells) or on day 6 (resting cells) after antigen stimulation, or isolated *ex vivo* from spinal cord parenchyma or CSF. Effector T cells from parenchyma were purified by a two layer OptiPrep (Axis-Shield PoC) density gradient. To block chemokine or integrin signalling T cells were incubated with either (1) PTX (0.2 μg ml⁻¹); (2) anti-rat monoclonal antibodies against VLA-4 (0.4 mg ml⁻¹) or LFA-1 (0.4 mg ml⁻¹); (3) mouse IgG2a (0.4 mg ml⁻¹) isotype control, or (4) cell culture medium alone for 1 h at 37 °C and washed thoroughly before injection. T_{MBP-GFP} or T_{OVA-Cherry}/T_{OVA-GFP} cells ($0.8–2 \times 10^6$ *ex vivo* cells or 5×10^6 *in vitro* cultured cells) were injected into the cisterna magna in a total volume of 40–50 μl . In some experiments T cells were injected by lumbar puncture at the level L3–L4 as described⁴⁴. Cells pre-incubated with antibody were tested by flow cytometry for antibody saturation before injection and after retrieval.

RNA extraction, cDNA library preparation and RNA-seq. RNA extraction, cDNA library preparation and RNA sequencing was undertaken as previously described⁴⁵. Total RNA was purified using the TRIzol protocol (Invitrogen) from 6 different samples: *in vitro* T_{MBP-GFP} cell blasts (20 h after antigen encounter), *in vitro* T_{MBP-GFP} resting T cells (day 6 after antigen encounter); *ex vivo* T_{MBP-GFP} cells from blood, CSF, spinal cord leptomeninges and parenchyma collected 3 days after transfer. Between 200,000–400,000 T_{MBP-GFP} cells were sorted from each sample with a BD FACSAria 4L SORP with more than 98% purity. Three different biological replicates were prepared for each sample. Five to seven animals were pooled for each *ex vivo* sample. Library preparation for RNA-seq was performed using the TruSeq RNA Sample Preparation Kit (Illumina) starting from 500 ng of total RNA. Single read (45 bp) sequencing was conducted using a HiSeq 2000 (Illumina). Fluorescence images were transformed to BCL files with the Illumina BaseCaller software. Samples were demultiplexed to FASTQ files with CASAVA. Sequencing quality was checked and approved via the FastQC software. Sequences were aligned to the genome reference sequence of *Rattus norvegicus* using the STAR alignment software⁴⁶ allowing for 2 mismatches within 45 bases. Subsequently, conversion of resulting SAM files to sorted BAM files, filtering of unique hits and counting was conducted with SAMtools⁴⁷ and HTSeq⁴⁸. Data was preprocessed and analysed in the R/Bioconductor environment using the DESeq2 package⁴⁹. Candidate genes were filtered to a minimum of twofold change and FDR-corrected P value < 0.05 . Gene annotation was performed using *Rattus norvegicus* entries from Ensembl v78 via the biomaRt package⁵⁰. KEGG pathway analysis was performed by using The Database for Annotation, Visualization and Integrated Discovery (DAVID)⁵¹.

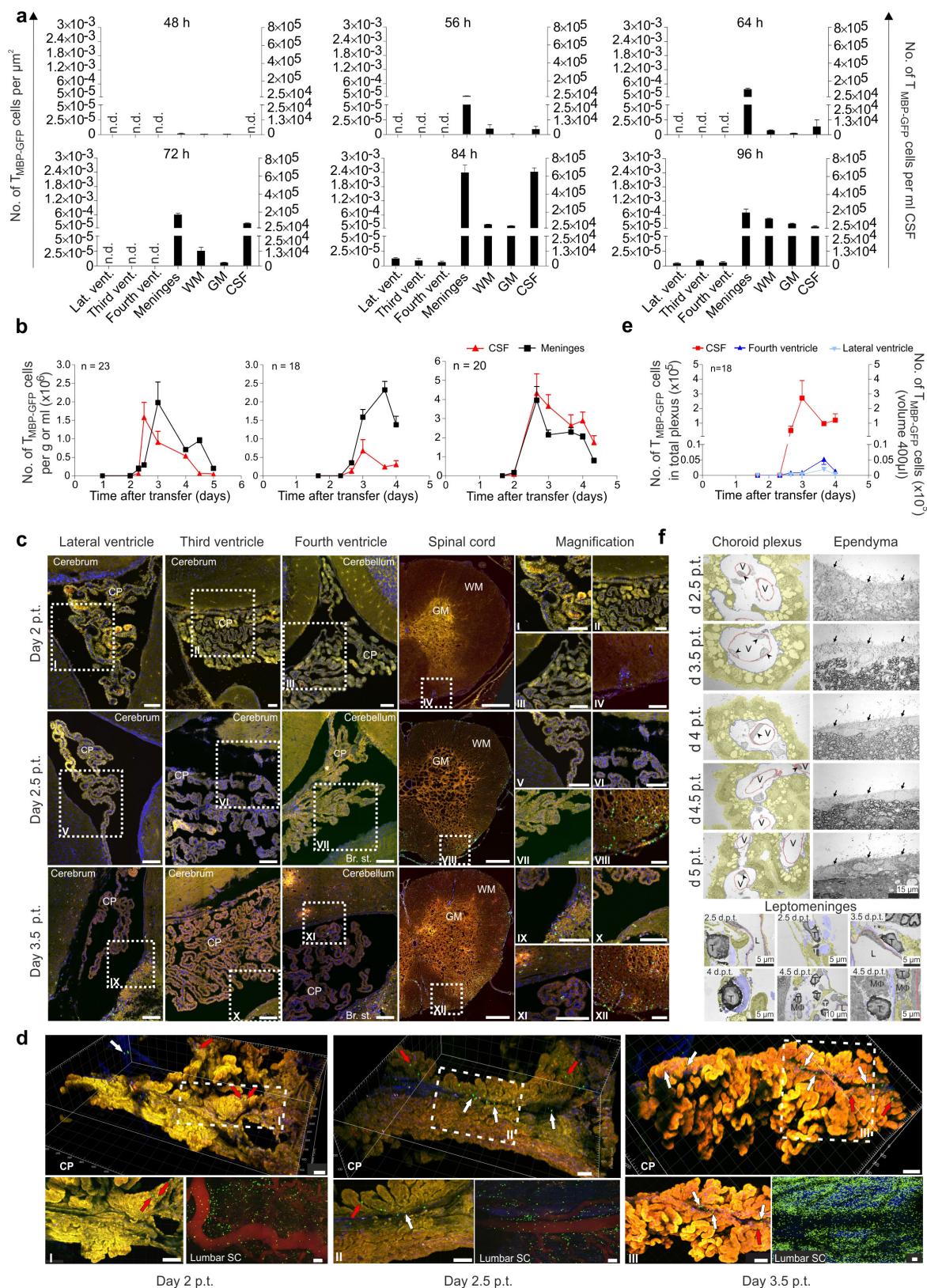
Quantitative PCR. mRNA extraction, cDNA synthesis and quantitative PCR were performed as described¹⁸. The following rat-specific primers and FAM-TAMRA-labelled probes not previously described^{7,18,52} were used. Vascular cell adhesion molecule 1 (VCAM-1): forward, 5'-TGAGGCTGGAATTAGCAAAAAAT-3', reverse, 5'-TTAGATGGAAAGACTGTAAGTTGTATG-3', probe: 5'-CTGATTATCCAAGGCTCTT-3'; intercellular adhesion molecule 1 (ICAM-1): forward, 5'-GAAGACAGCAGACCACTGTGCTT-3', reverse, 5'-CTCGCTCTGGGGAA CGAATACA-3', probe: 5'-ACTGTGGCACCGC-3'; fibronectin transcript-1 (fn-1): forward, 5'-TGATCTTGAGGAACATGGCTTT-3', reverse, 5'-GCAGGT ATGGTCTGGCCTAAG-3', probe, 5'-AACACGCCACCCACTGCGG-3'; fibronectin transcript-2 (fn-1): forward, 5'-GAGCTTCCCAACTGGTTAAC-3', reverse, 5'-GAACTGTGGAGGGAACATCCAA-3', probe, 5'-CCACACCCCAA TCTTCATGGACCAGA-3'. Unless otherwise stated, β -actin was used as house-keeping gene.

In vitro proliferation assay of T cells. Amplification of *ex vivo* or *in vitro* isolated T_{MBP}-GFP cells upon antigen stimulation was tested by flow cytometry as previously described¹⁸.

Histology and immunohistochemistry. Histological analysis was performed as described³⁴ using the following antibodies: rabbit anti-CXCL-11 (Biorbyt), mouse anti-rat RECA-1 (Abcam), guinea-pig anti-Iba-1 (Synaptic Systems), mouse anti-rat OX-6 (rat RT1B MHC class II antigen, Serotec), mouse anti-alpha-smooth muscle actin (DAKO). eGFP was enhanced by GFP-Atto 488 (Chromotek). Alexa Fluor568 goat anti-rabbit IgG, Alexa Fluor647 goat anti-mouse IgG (both from Invitrogen), Alexa Fluor647 donkey anti-guinea pig (Dianova) were used as control. Images were acquired using either a Zeiss Axio Observer fluorescence Microscope equipped with a 10 \times Air Objective or a Zeiss LSM700 confocal microscope equipped with a 40 \times or 63 \times Zeiss objective. For histological quantification of T_{MBP}-GFP cells in CNS areas, consecutive cryosections (16 μ m) were acquired with a VS120 Virtual Slide Microscope (Olympus) equipped with a 10x objective. Counting of T_{MBP}-GFP cells and area measurements were performed with Fiji software⁵³. For CNS explant imaging, lumbar spinal cord and fourth ventricle choroid plexus were prepared. Images were acquired by TPLSM. For detection of integrin ligands *in vivo* on leptomeningeal structures dura mater and arachnoidea were removed leaving the subarachnoidal structures intact. PE-labelled VCAM or ICAM antibodies (BioLegend) and the respective isotype controls were then incubated *in situ* for 30 min before confocal acquisition.

Electron Microscopy. Animals were perfused with saline followed by a fixative containing 4% PFA and 1% glutaraldehyde for ultrastructural analysis and 4% PFA and 0.1% glutaraldehyde for immuno-electron microscopy. After post-fixation the spinal cord was cut into 60 μ m sections using a Leica vibrating microtome. For immuno-electron microscopy, goat anti-GFP antibody (Acris antibodies) was used to detect the GFP signal. Biotinylated rabbit anti-goat antibody (Sigma Aldrich) was used as secondary antibody. The sections were then incubated with ExtrAvidin (Sigma-Aldrich) and stained with diaminobenzidine (DAB, Sigma-Aldrich) to achieve an electron dense precipitate. Omitting primary antibodies resulted in the absence of staining. After that, the sections were stained in 0.5% osmium tetroxide, dehydrated and embedded in durcupan (Sigma-Aldrich). Regions of interest were identified by light microscopy, cut and transferred onto blocks of resin to obtain ultra-thin sections using an Ultramicrotome (Leica). Ultra-thin sections were transferred on formvar-coated copper grids and stained using lead citrate. Ultrastructural analysis was performed using a Zeiss SIGMA electron microscope equipped with a STEM detector and ATLAS software.

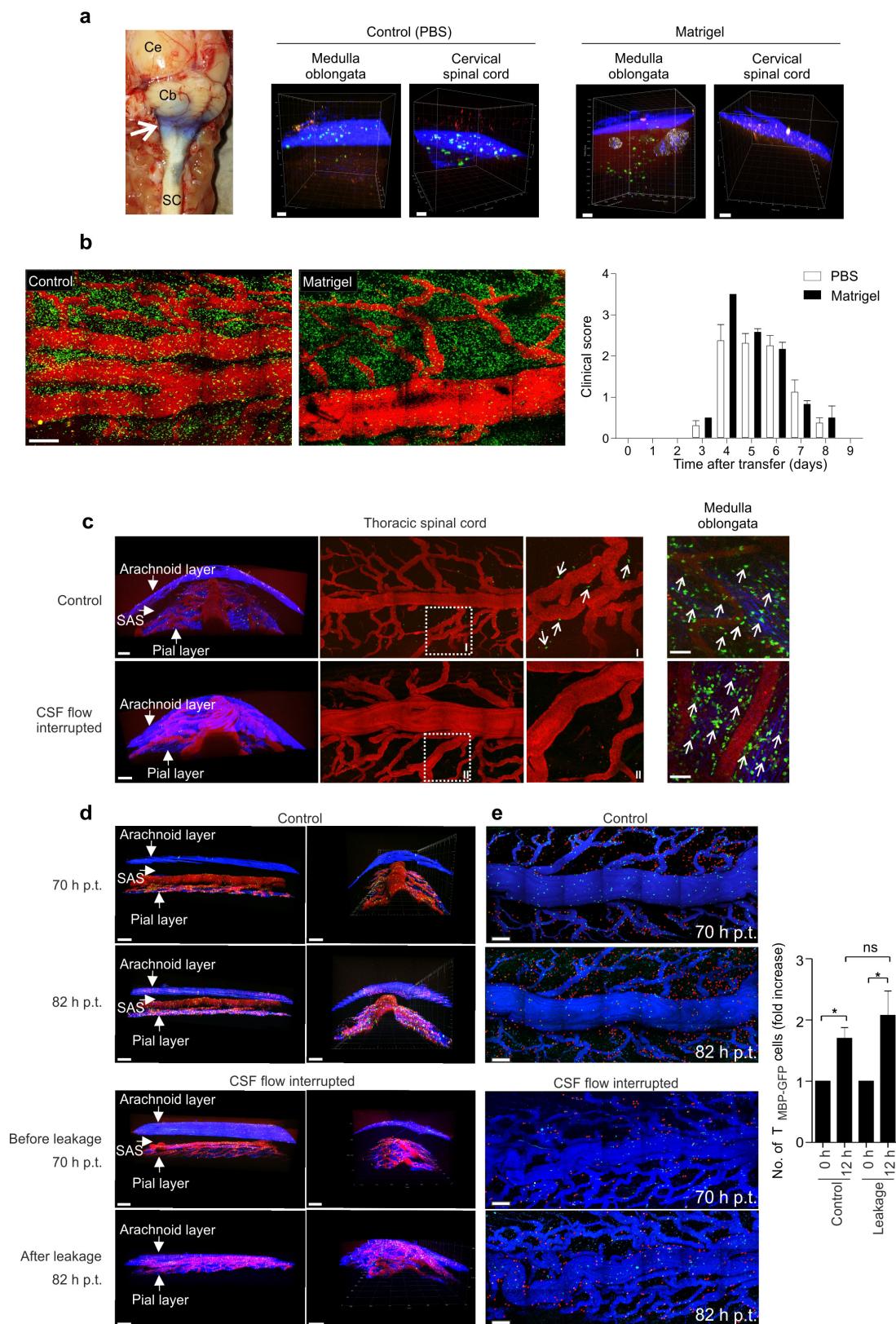
30. Flügel, A., Willem, M., Berkowicz, T. & Wekerle, H. Gene transfer into CD4 $^{+}$ T lymphocytes: green fluorescent protein-engineered, encephalitogenic T cells illuminate brain autoimmune responses. *Nature Med.* **5**, 843–847 (1999).
31. Gurskaya, N. G. *et al.* Engineering of a monomeric green-to-red photoactivatable fluorescent protein induced by blue light. *Nature Biotechnol.* **24**, 461–465 (2006).
32. Eylar, E. H., Kniskern, P. J. & Jackson, J. J. Myelin basic proteins. *Methods Enzymol.* **32**, 323–341 (1974).
33. Adelmann, M. *et al.* The N-terminal domain of the myelin oligodendrocyte glycoprotein (MOG) induces acute demyelinating experimental autoimmune encephalomyelitis in the Lewis rat. *J. Neuroimmunol.* **63**, 17–27 (1995).
34. Flügel, A. *et al.* Migratory activity and functional changes of green fluorescent effector cells before and during experimental autoimmune encephalomyelitis. *Immunity* **14**, 547–560 (2001).
35. Luque, A. *et al.* Activated conformations of very late activation integrins detected by a group of antibodies (HUTS) specific for a novel regulatory region (355–425) of the common $\beta 1$ chain. *J. Biol. Chem.* **271**, 11067–11075 (1996).
36. Siffrin, V. *et al.* Differential immune cell dynamics in the CNS cause CD4 $^{+}$ T cell compartmentalization. *Brain* **132**, 1247–1258 (2009).
37. Kawakami, N. *et al.* Live imaging of effector cell trafficking and autoantigen recognition within the unfolding autoimmune encephalomyelitis lesion. *J. Exp. Med.* **201**, 1805–1814 (2005).
38. Issekutz, T. B. Inhibition of *in vivo* lymphocyte migration to inflammation and homing to lymphoid tissues by the TA-2 monoclonal antibody. A likely role for VLA-4 *in vivo*. *J. Immunol.* **147**, 4178–4184 (1991).
39. Sporici, R. & Issekutz, T. B. CXCR3 blockade inhibits T-cell migration into the CNS during EAE and prevents development of adoptively transferred, but not actively induced, disease. *Eur. J. Immunol.* **40**, 2751–2761 (2010).
40. Matsui, M. *et al.* Treatment of experimental autoimmune encephalomyelitis with the chemokine receptor antagonist Met-RANTES. *J. Neuroimmunol.* **128**, 16–22 (2002).
41. Meiron, M., Zohar, Y., Anunu, R., Wildbaum, G. & Karin, N. CXCL12 (SDF-1alpha) suppresses ongoing experimental autoimmune encephalomyelitis by selecting antigen-specific regulatory T cells. *J. Exp. Med.* **205**, 2643–2655 (2008).
42. Carnevali, L. *et al.* Vagal withdrawal and susceptibility to cardiac arrhythmias in rats with high trait aggressiveness. *PLoS ONE* **8**, e68316 (2013).
43. Slobodian, I., Krassioukov-Enns, D. & Del Bigio, M. R. Protein and synthetic polymer injection for induction of obstructive hydrocephalus in rats. *Cerebrospinal Fluid Res.* **4**, 9 (2007).
44. De la Calle, J. L. & Paineo, C. L. A procedure for direct lumbar puncture in rats. *Brain Res. Bull.* **59**, 245–250 (2002).
45. Cabeza, R. *et al.* An RNA sequencing transcriptome analysis reveals novel insights into molecular aspects of the nitrate impact on the nodule activity of Medicago truncatula. *Plant Physiol.* **164**, 400–411 (2014).
46. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
47. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
48. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
49. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
50. Durinck, S. *et al.* BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* **21**, 3439–3440 (2005).
51. Huang, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**, 44–57 (2009).
52. Odoardi, F. *et al.* Instant effect of soluble antigen on effector T cells in peripheral immune organs during immunotherapy of autoimmune encephalomyelitis. *Proc. Natl. Acad. Sci. USA* **104**, 920–925 (2007).
53. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**, 676–682 (2012).



Extended Data Figure 1 | See next page for figure caption.

Extended Data Figure 1 | T_{MBP} cells enter the CSF from the leptomeninges. **a**, T_{MBP} cell entry into the CNS compartments. $T_{MBP-GFP}$ cells were quantified by histological analysis in the choroid plexus of all four ventricles (1st–4th vent) and spinal cord (meninges, white matter (WM) and grey matter (GM)) and by flow cytometry in the CSF. Histology of 6–27 consecutive slices per compartment. n.d., not detectable. Representative experiment \pm s.e.m. from 2 independent experiments ($n = 12$) from days 1.5–5 after transfer. **b**, Accumulation of $T_{MBP-GFP}$ cells in the CSF and leptomeninges occurs simultaneously. Numbers of $T_{MBP-GFP}$ cells (flow cytometry) in meninges and CSF. Results of 3 independent experiments. Data are mean \pm s.d. of 2–4 animals per time point ($n = 61$). **c**, T_{MBP} cells accumulate in the leptomeninges but not in the choroid plexus. Confocal laser scanning microscopy of fixed tissue sections of the lateral, third and fourth ventricles and fluorescence microscopy images of lumbar spinal cord sections. Images were recorded from the same animal for each time point. Br. st., brain stem; green, $T_{MBP-GFP}$ cells; blue, DAPI counterstain of nuclei. Right (I–XII), magnifications of areas of interest. Representative images of 14 different time points from days 1.5–5 after T-cell transfer. **d**, Localization of T_{MBP} cells in choroid plexus explants compared to spinal cord leptomeninges. Original TPLSM images and 3D reconstructions of explanted choroid plexus of the fourth ventricle. The corresponding lumbar spinal cord leptomeninges (lumbar SC) was acquired before explantation of the choroid plexus.

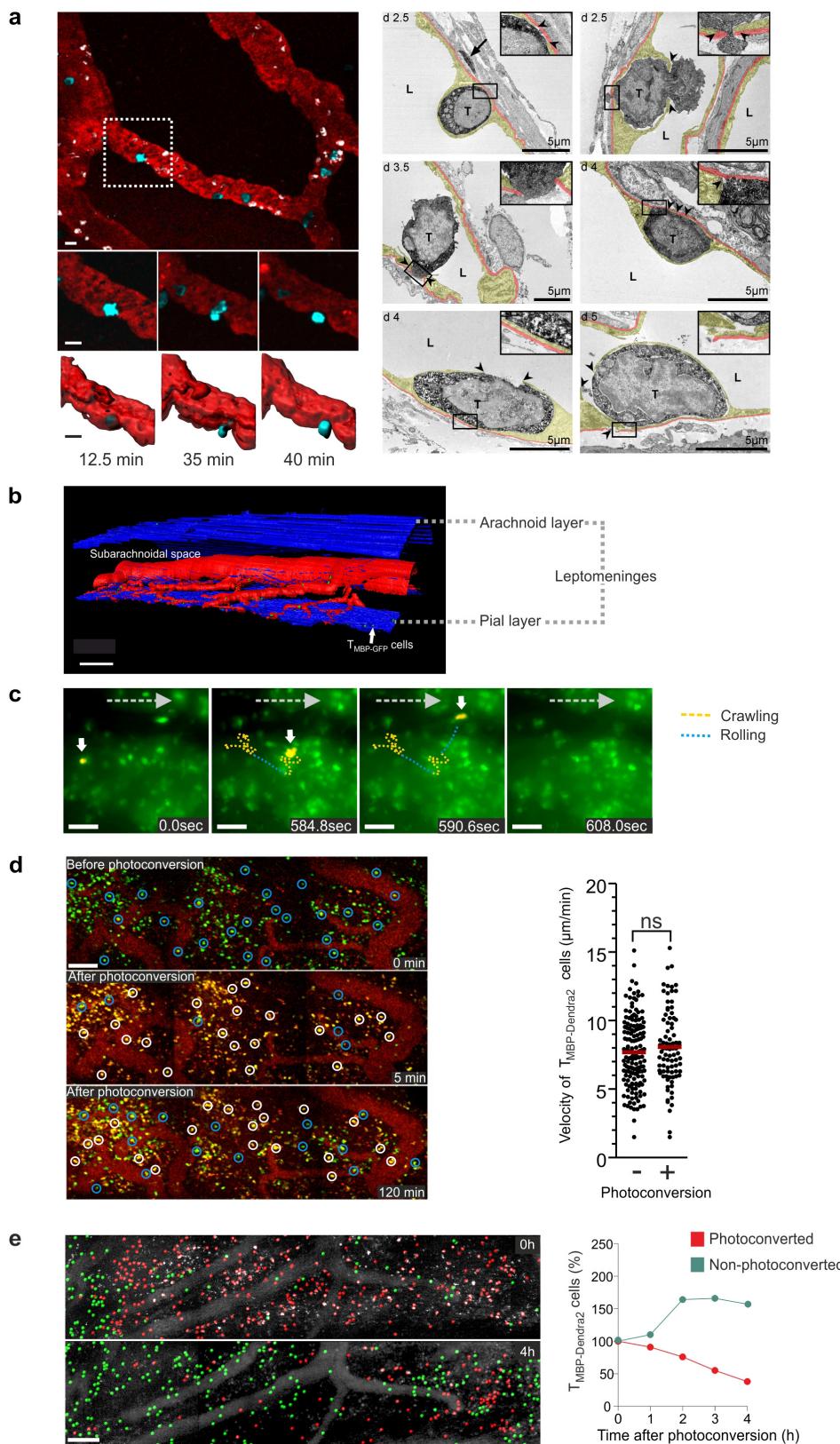
Green, $T_{MBP-GFP}$ cells; red, meningeal blood vessels; blue, collagen. Few $T_{MBP-GFP}$ cells can be detected within the choroid plexus tissues compared to the leptomeninges of the spinal cord. Red arrows, $T_{MBP-GFP}$ cells within the choroid plexus stroma; white arrows, $T_{MBP-GFP}$ cells outside of the choroid plexus tissue. Representative pictures of 2 independent experiments ($n = 10$). Scale bars, 100 μ m (**c, d**). **e**, T_{MBP} cells accumulate in the CSF before appearing in the choroid plexus. Quantification of $T_{MBP-GFP}$ cells from the choroid plexus of the fourth and lateral ventricles and from the CSF (flow cytometry). Means \pm s.d. Representative data of 3 independent experiments including 2–4 animals per time point. **f**, Anti-GFP immuno-electron microscopy confirms the scarcity of T_{MBP} cells in the choroid plexus. Upper left, choroid plexus of the fourth ventricle. The choroid plexus epithelial cells are colour-coded in yellow. The lamina propria between epithelial basal lamina (blue) and vascular basal lamina (red) is filled with loose connective tissue and meningeal fibroblasts. Arrow heads, endothelial nuclei of cross sectioned vessels (V). Top right, ependyma. The surface of the ventricular ependyma (arrows) appears smooth and continuous with numerous ciliary processes. No T_{MBP} cells could be detected for any of the observed time points after T-cell transfer. Lower panel, leptomeninges. $T_{MBP-GFP}$ cells (T) are marked by typical black grains of DAB. Yellow, resident leptomeningeal cells; blue, bundles of collagen fibres; red, vascular endothelial basal laminas. M Φ , macrophages; L, vascular lumen.



Extended Data Figure 2 | See next page for figure caption.

Extended Data Figure 2 | Interference with T_{MBP} cell transport within the CSF does not inhibit their accumulation in the spinal cord and clinical EAE. **a**, T_{MBP} cell transport from the cisterna magna to the subarachnoidal space of the spinal cord is effectively blocked by Matrigel. Left, macroscopic image of the cisterna magna and the adjacent parts of the CNS after injection of Matrigel mixed with Evans blue dye. Cb, cerebellum; Ce, cerebrum; SC, spinal cord. Arrow, cisterna magna filled with blue-stained Matrigel. Right, PBS or Matrigel was injected in the cisterna magna of naive animals. $T_{MBP-GFP}$ cells were injected 24 h later intra-cisternally. TPLSM of the medulla oblongata and the cervical spinal cord was performed 6 h after the $T_{MBP-GFP}$ cell injection. Shown are 3D reconstructions of the TPLSM recordings. Notably, in Matrigel-injected animals the T_{MBP} cells remained localized in the cisterna magna but did not reach the cervical spinal cord, indicating that the Matrigel efficiently blocked the migration of cells from the cisterna magna to the subjacent leptomeninges. In controls (PBS-injected animals), T_{MBP} cells readily reached the leptomeninges of the cervical spinal cord. Green, $T_{MBP-GFP}$ cells; blue, collagen. Scale bars, 50 μ m. **b**, Leptomeningeal T-cell infiltration and clinical disease are not impaired after Matrigel blockage. Matrigel was injected into the cisterna magna 2 days after i.v. transfer of $T_{MBP-GFP}$ cells. PBS i.t.-injected animals were used as control. Left, TPLSM recordings of thoracic leptomeninges 3.5 days after transfer. Representative data of 2 independent experiments. Scale bar, 200 μ m. Right, clinical score assessment. Representative data of 2 independent experiments including 4 animals per treatment ($n = 16$). Mean \pm s.e.m. **c**, T_{MBP} cell transport from the cisterna magna to the subarachnoidal space of the spinal cord is effectively blocked by interrupting the CSF flow. Resting T cells (5×10^6) were injected into the cisterna magna either in control animals (upper panel) or in animals where the CSF flow was interrupted by introducing a CSF leakage at the level of the cervical spinal cord (lower panel). The efflux of the CSF induced a collapse of the

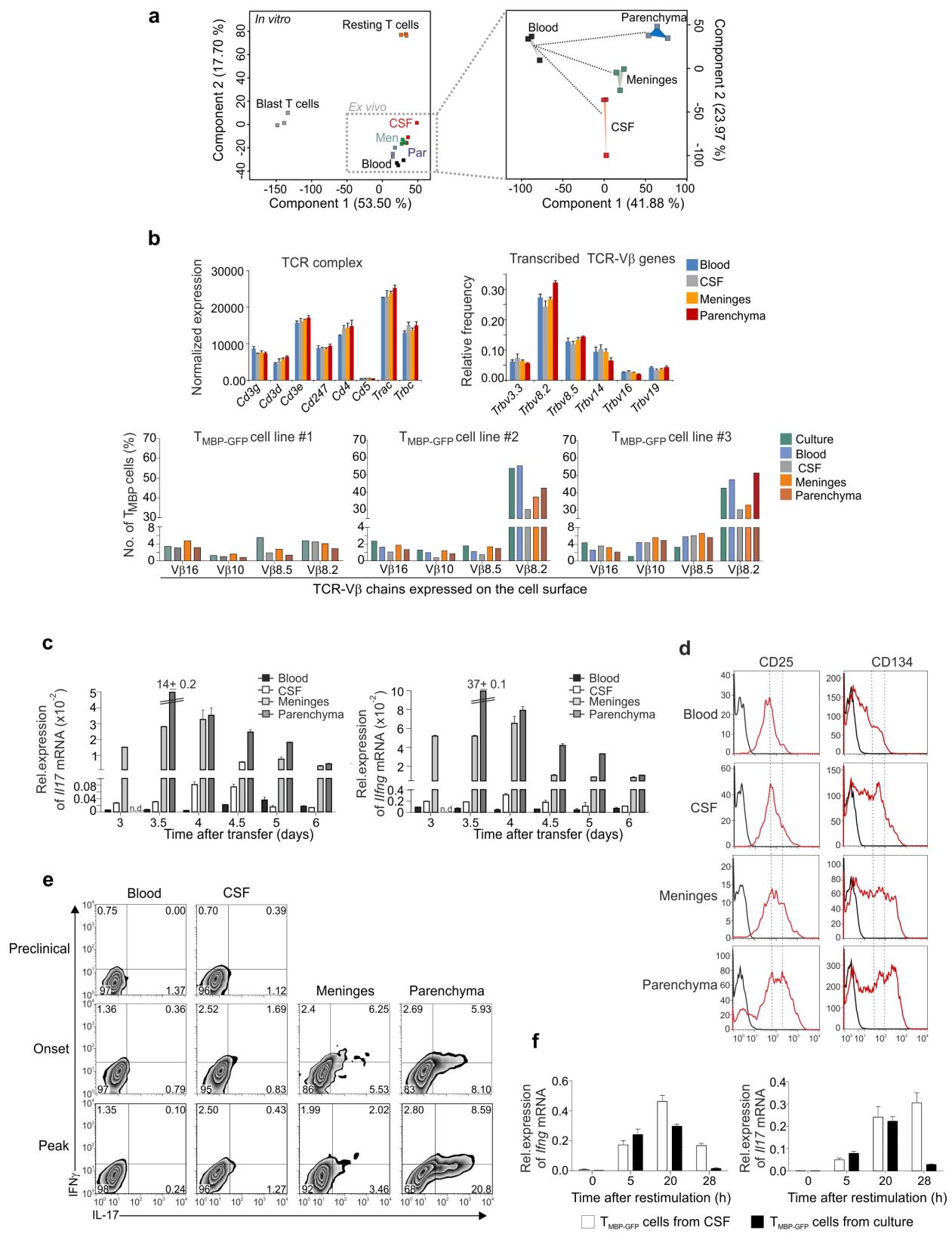
subjacent thoracic subarachnoidal space (SAS) and thereby abolished the CSF flow in this area. TPLSM recordings of the thoracic spinal cord and of the medulla oblongata 6 h after i.t. injection. Depicted are 3D reconstructions of TPLSM recordings of the thoracic spinal cord leptomeninges, overviews of the same area with magnifications of areas of interest (I, II) and overviews of the medulla oblongata. Green, $T_{MBP-GFP}$ cells; red, blood vessels; blue, collagen. Arrows, representative $T_{MBP-GFP}$ cells. Scale bars, 300 μ m for thoracic spinal cord and 50 μ m for medulla oblongata. Note that in the animals with disrupted CSF flow, no T_{MBP} cells were detectable at the level of the spinal cord despite the high number of T_{MBP} cells in the medulla oblongata indicating that the leakage completely blocked the transport of cells from the cisterna magna to the subjacent spinal cord leptomeninges. **d, e**, Interrupting the CSF flow does not impair T_{MBP} cell accumulation in the spinal cord leptomeninges. **d**, Shown are 3D reconstructions of TPLSM recordings of the thoracic spinal cord leptomeninges after i.v. transfer of $T_{MBP-GFP}$ cells. Images before (70 h after transfer) and after introducing a CSF leakage at the level of the cervical spinal cord (82 h after transfer). Animals with intact flow were used as control. Green, $T_{MBP-GFP}$ cells; red, meningeal blood vessels; blue, collagen. Scale bars, 300 μ m. **e**, Left, TPLSM overviews of the thoracic spinal cord leptomeninges. The numbers of $T_{MBP-GFP}$ cells that had emigrated from the local vessels into the leptomeningeal milieu (red false colours) within the 12 h interval did not differ between controls and 'CSF flow interrupted' animals. Green and red marks (overlaid red dots), T_{MBP} cells located inside or outside the vessels, respectively; blue (false colour), meningeal blood vessels. Scale bars, 200 μ m. Right, relative increase of $T_{MBP-GFP}$ cells located in the extravascular compartment at the indicated time points during imaging either in control animals or in animals where the CSF was interrupted (leakage). Representative data of 3 independent experiments (Kruskal–Wallis ANOVA followed by Dunn's multiple comparison test, $*P < 0.05$). Mean \pm s.d.



Extended Data Figure 3 | See next page for figure caption.

Extended Data Figure 3 | T_{MBP} cells entering the leptomeningeal milieu from local vessels become detached and float within the CSF. **a**, T_{MBP} cell extravasation from the leptomeningeal vessels. Left, TPLSM during the pre-invasion phase. Representative T_{MBP} cell diapedesis. Depicted is a very early transmigration event, that is, the majority of both T_{MBP} cells and $CD11b^+$ myeloid cells are located inside the vascular lumina. Red, meningeal blood vessels; blue, T_{MBP} -Lifeact-Turquoise2 cells; white, $CD11b^+$ myeloid cells labelled by i.v. administration of fluorescently conjugated anti- $CD11b$ monoclonal antibody during imaging. Scale bars, 10 μ m. Right, immuno-electron microscopy of T_{MBP} cell extravasations observed at the indicated time points after transfer. Yellow, endothelial layer of leptomeningeal vessels; red, endothelial basal lamina. T_{MBP} -GFP cells are marked by cytoplasmic DAB. Extravasating T_{MBP} cells were often found to be surrounded by endothelial processes. Thereby, notable abluminal or luminal gaps (arrow heads) of the endothelial layer were frequently observed. Representative electron micrographs at the indicated time points. L, vascular lumen; T, T cells. **b**, Surface-rendered 3D reconstruction of the meningeal compartment. TPLSM recording in the early phase of leptomeningeal T_{MBP} cell infiltration. Red, meningeal blood vessels; green, T_{MBP} -GFP cells; blue, collagen. Scale bar, 300 μ m. **c**, Visualization of a T_{MBP} cell detaching from the leptomeninges into the CSF. Intravital video-microscopy during the established phase of leptomeningeal T_{MBP} cell infiltration depicts a T_{MBP} -GFP cell (false colour

yellow, indicated by the white arrow) in the process of detachment from the pial surface. Yellow and blue lines, crawling and rolling steps, respectively; dotted grey arrows, direction of the CSF flow. Scale bars, 50 μ m. **d**, *In situ* light exposure induces efficient and high contrast photoconversion of T_{MBP} -Dendra2 cells without influencing T-cell motility. Images show representative TPLSM leptomeningeal overviews during established leptomeningeal T_{MBP} cell infiltration before (0 min), 5 min, and 120 min after photoconversion. Red, meningeal blood vessels; green, non-photoconverted T_{MBP} -Dendra2 cells; yellow, photoconverted T_{MBP} -Dendra2 cells; blue and white circles, representative non-photoconverted or photoconverted T_{MBP} -Dendra2 cells, respectively. Plotted are mean velocities of individual T_{MBP} -Dendra2 cells before and 2 h after photoconversion. Data refer to 30 min time-lapse recordings. Mean values from 239 cells from at least 3 independent experiments (two-tailed Mann–Whitney *U*-test). **e**, Rapid turnover of effector T cells in the leptomeninges. Intravital TPLSM recordings on leptomeninges during the established phase of leptomeningeal T_{MBP} cell infiltration. Images show distribution of T_{MBP} -Dendra2 cells before or 4 h after photoconversion. For better visualization, red and green dots were overlaid onto photoconverted and non-photoconverted T cells, respectively. Grey (false colour), vessel lumen. Scale bar, 100 μ m. Right, relative changes of photoconverted versus non-photoconverted T_{MBP} -Dendra2 cells over time obtained from intravital recordings. Representative data of at least 3 independent experiments.



Extended Data Figure 4 | See next page for figure caption.

Extended Data Figure 4 | CSF-derived T_{MBP} cells do not represent a specialized sub-population compared to T_{MBP} cells in the meninges and CNS parenchyma but they display a lower activation profile.

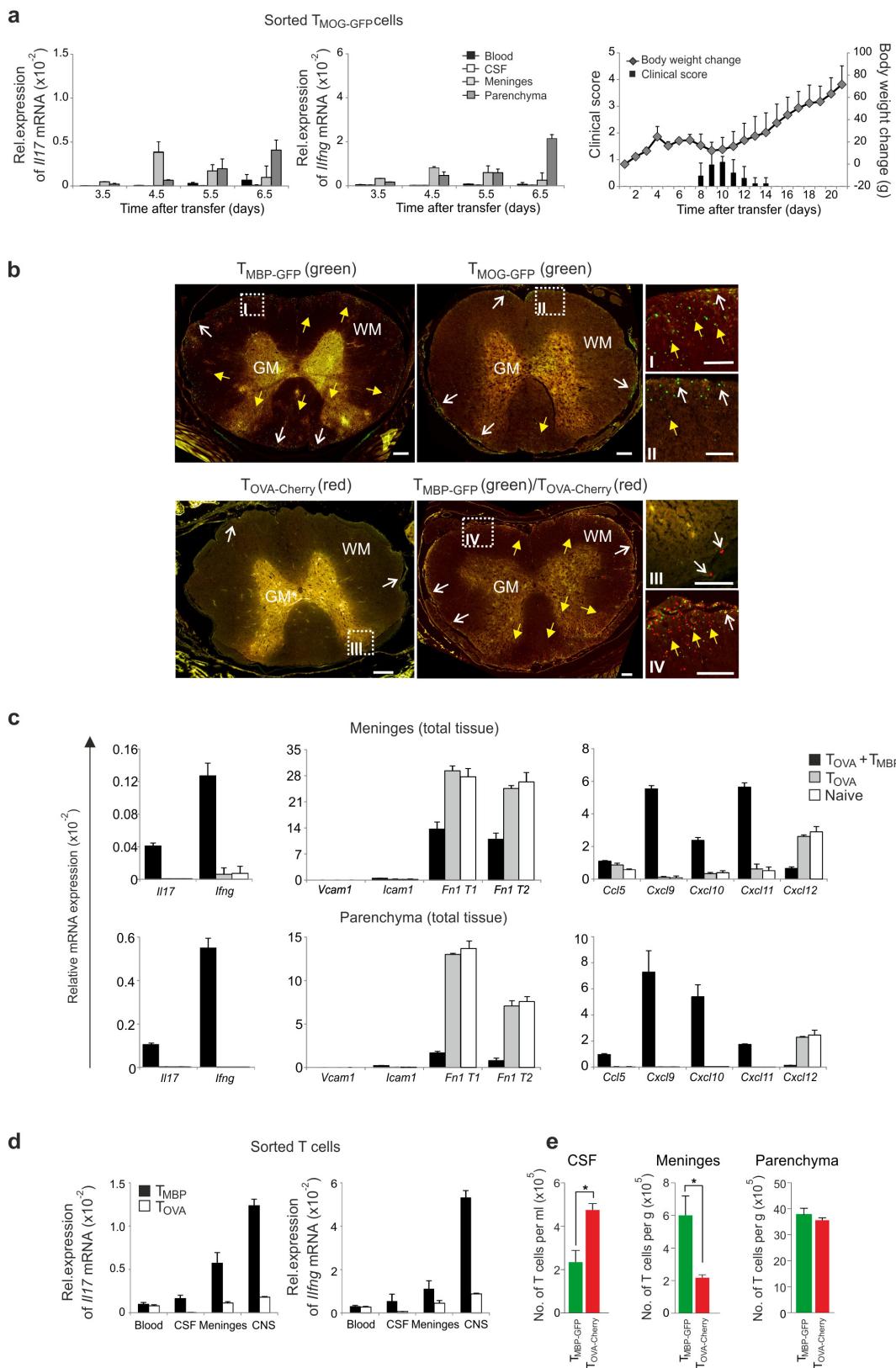
a, Similarities in transcriptome profiles between T_{MBP} cells in blood, CSF, leptomeninges and CNS parenchyma. RNA-seq of $T_{MBP-GFP}$ cells sorted from blood, meninges (Men), CSF and CNS parenchyma (Par) 3 days after transfer and on *in vitro* cultured $T_{MBP-GFP}$ cells sorted 20 h and 6 days after antigenic stimulation (blast and resting T cells, respectively). Principal component analysis of the transcriptomes for all six T cell populations (left) and for the four *ex vivo* populations (right) show similar profiles of $T_{MBP-GFP}$ cells in the CSF and the other CNS compartments compared to blood and culture. Numbers in parentheses, proportion of total variability calculated for each principal component. Each data point represents a biological replicate. **b**, T_{MBP} cells display similar TCR repertoires in the CNS compartments. Top, RNA-seq as in **a**. Normalized expression of invariant TCR complex components (left) and mean frequencies of the TCR $V\beta$ genes (right) determined as the proportion of RNA-seq reads that map to certain $V\beta$ segments among all reads mapping to the entire set of 24 rat $V\beta$ segments. Shown are the 6 most abundant $V\beta$ segments. Bottom, $T_{MBP-GFP}$ cells from culture (4 days after antigen stimulation), blood, or from the indicated CNS compartments on day 4 after T-cell transfer. Expression (flow cytometry) of different TCR $V\beta$ chains. Percentages of the different $V\beta$ chains for three different T_{MBP} cell lines.

Ex vivo cell staining containing pooled cells from the respective organs of 3–4 animals ($n=10$). **c–e**, T_{MBP} cells within blood and CSF are not activated in contrast to T_{MBP} cells in meninges and CNS parenchyma.

c, Amounts of IFN γ and IL-17 (quantitative PCR) in T_{MBP} cells from blood or the indicated CNS compartments at the indicated time points after transfer. Data are mean \pm s.d. of duplicate measurements. Representative data of 4 independent experiments including at least 2 animals per group per time point. n.d., not determined owing to lack of cells within the CNS parenchyma at the early leptomeningeal T_{MBP} cell infiltration phase.

d, Cell surface expression of activation markers CD25 and CD134 (flow cytometry) of $T_{MBP-GFP}$ cells at day 4 after transfer. Black, isotype control; red, CD25 or CD134. Representative data of 4 independent experiments each combining cells from 4 or 5 animals for each compartment.

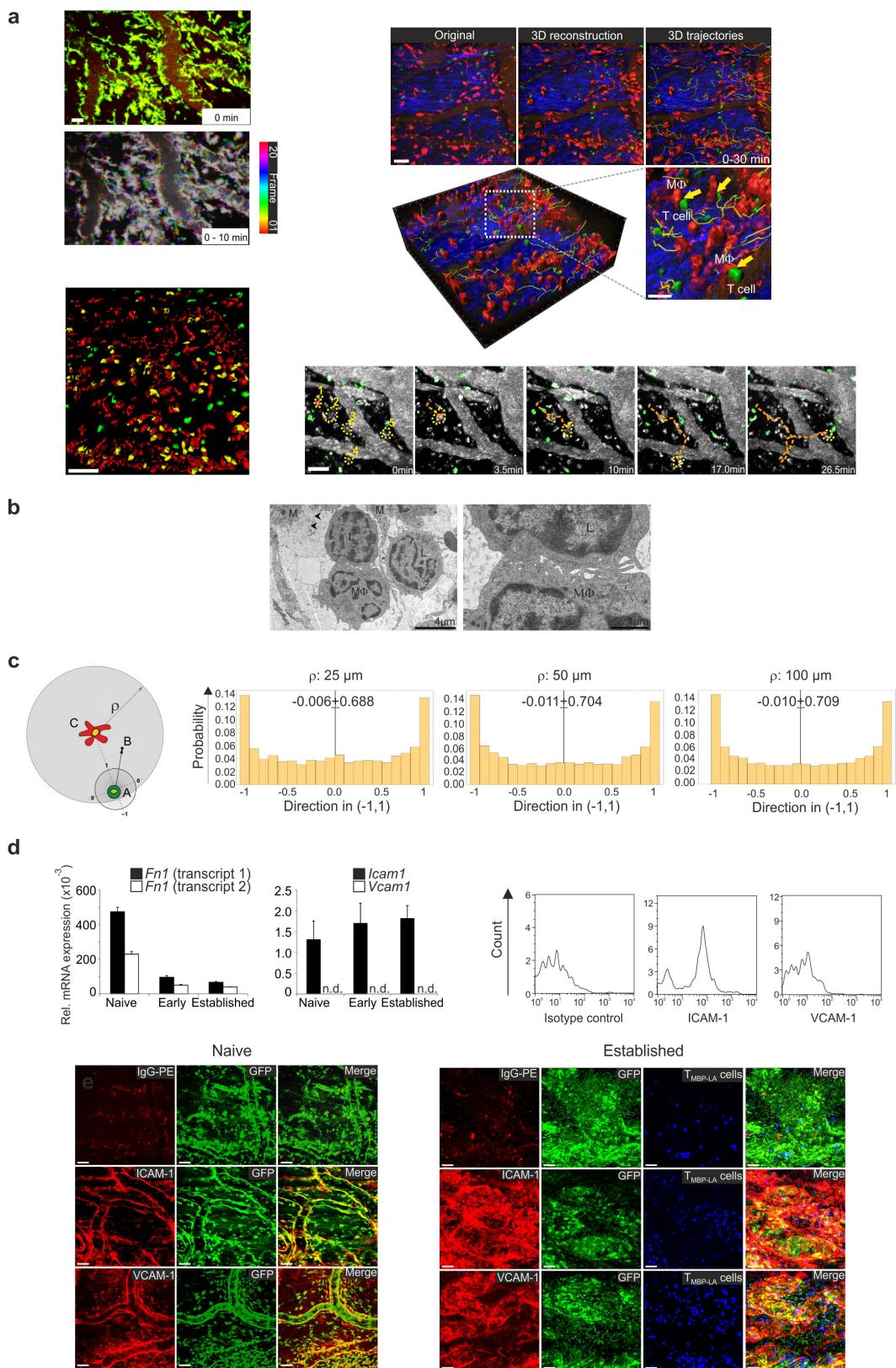
e, Intracellular IFN γ and IL-17 production in T_{MBP} cells isolated during the indicated phases of EAE. Representative data of 4 independent experiments each combining cells from 4 or 5 animals for each compartment ($n=17$). **f**, CSF-derived T_{MBP} cells produce cytokines upon antigenic stimulation in CSF. $T_{MBP-GFP}$ cells isolated at day 3 after transfer from CSF or cultured resting $T_{MBP-GFP}$ cells were stimulated *in vitro* with MBP in CSF from naive rats. IFN γ and IL-17 production (quantitative PCR). Data are mean \pm s.d. of duplicate measurements representative of 2 independent experiments ($n=8$). House-keeping gene, β -actin (*Actb*) (**c, f**).



Extended Data Figure 5 | See next page for figure caption.

Extended Data Figure 5 | The trafficking of effector T cells in the CNS compartments depends on their reactivation levels and the inflammatory state in the CNS. **a**, T_{MOG} cells have a low encephalitogenic potential and display low reactivation levels in the CNS. Left, IL-17 and IFN γ expression measured by quantitative PCR in $T_{MOG-GFP}$ cells. House-keeping gene, β -actin (*Actb*) (**a, c, d**). Representative data of 3 independent experiments (data are mean \pm s.d. of duplicate measurements). Note, T_{MOG} cells in meninges and CNS parenchyma produce less cytokines than the highly pathogenic T_{MBP} cells (Extended Data Fig. 4c), but more than non-pathogenic T_{OVA} cells (Extended Data Fig. 4d). Right, clinical course after $T_{MOG-GFP}$ cell i.v. transfer. Cumulative data of 3 independent experiments including at least 4 animals per group. T_{MOG} cell CNS infiltration in this model considerably precedes disease onset (Fig. 2e). **b**, Morphological analysis of the distribution of effector T cells with different antigen specificities and reactivation/pathogenic potentials in the CNS. Fluorescence-microscopy-derived overviews and magnified subsets (I–IV) of thoracic spinal cord sections showing distribution of fluorescently labelled T cells with the indicated antigen specificity 4 days after T-cell transfer. Open white and closed yellow arrows, representative T cells located in the meningeal compartment or in the CNS parenchyma, respectively. Scale bars: overviews, 200 μ m; magnification, 100 μ m.

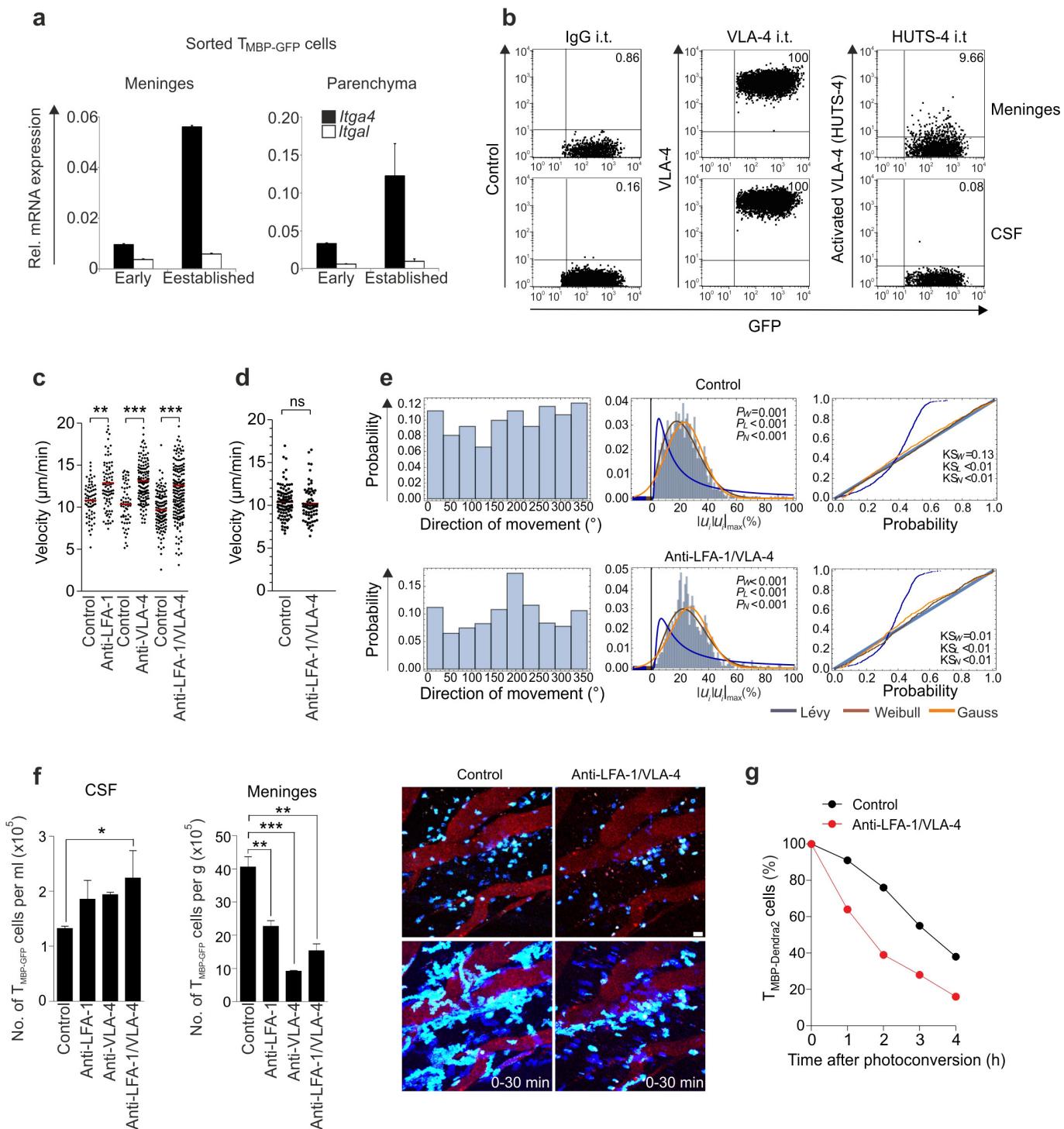
The infiltration behaviour of brain-non-reactive T_{OVA} cells is similar to that of pathogenic T_{MBP} cells when the CNS tissue is inflamed (after co-transfer of the T_{OVA} cells with T_{MBP} cells). **c**, T_{OVA} cells encounter an inflammatory milieu in the meninges when co-transferred with T_{MBP} cells. Quantitative PCR for the indicated chemokines, integrin ligands and cytokines was performed on meninges and CNS parenchyma isolated from naïve animals (white columns) or from animals that 4 days previously had received either $T_{OVA-Cherry}$ cells (grey columns) or $T_{OVA-Cherry}$ cells together with $T_{MBP-GFP}$ cells (black columns). **d**, When co-transferred with T_{MBP} cells, T_{OVA} cells do not show signs of activation in the CNS. IFN γ and IL-17 expression (quantitative PCR) in $T_{MBP-GFP}$ or $T_{OVA-Cherry}$ cells sorted from the indicated compartments in animals transferred 4 days previously with these cells. Representative data of 3 independent experiments; data are mean \pm s.d. of duplicate measurements (**c, d**). **e**, T_{OVA} cells are released from the meningeal compartment into the CSF in higher numbers than T_{MBP} cells. Quantification of $T_{OVA-Cherry}$ cells and co-injected $T_{MBP-GFP}$ cells during established leptomeningeal T_{MBP} cell infiltration (flow cytometry). Representative data of 3 independent experiments including 3 animals per group. Data are mean \pm s.e.m. (two-tailed Mann–Whitney *U*-test).



Extended Data Figure 6 | See next page for figure caption.

Extended Data Figure 6 | T_{MBP} cells in the leptomeningeal milieu closely interact with resident macrophages. **a**, Composition of the leptomeningeal milieu and interactions of T_{MBP} cells with resident macrophages. Upper left, representative still and time-encoded projection illustrating, via a time colour-code scale, the area that single leptomeningeal macrophages scan with their processes over a recording period of 10 min (32 s time intervals). Green, GFP^+ meningeal macrophages in GFP^+ bone marrow chimaeras⁷; red, blood vessels. Upper right, original TPLSM picture and surface-rendered 3D reconstructions of a representative spot within the meningeal milieu during established leptomeningeal T_{MBP} cell infiltration. Green, extravasated motile $T_{MBP-GFP}$ cells; red, meningeal phagocytes ($M\Phi$); yellow trajectories, migration paths of individual $T_{MBP-GFP}$ cells; blue, collagen fibres. Magnification of an individual region (white dotted rectangle) indicates contacts (yellow arrows) between $T_{MBP-GFP}$ cells and meningeal phagocytes. Lower left, T_{MBP} cells in direct contact (yellow) or not in contact (green) with meningeal macrophages (red). Surface-rendered 3D reconstruction of a representative still from a 30 min TPLSM recording. Lower right, a representative $T_{MBP-GFP}$ effector cell (red, false colour) migrating within the leptomeningeal compartment contacting several resident meningeal macrophages (1–5, yellow contours). Orange line, migration path of the T cell; green, $T_{MBP-GFP}$ cells; grey, vessel lumen (false colour). Scale bars, 50 μ m. **b**, Ultrastructural analysis confirms direct and intense contacts between T cells and meningeal macrophages. Representative overview of cells within the spinal cord leptomeninges illustrating the organization of the leptomeningeal milieu at the ultrastructural level. Lymphocytes (L) in the subarachnoidal space are located between or adjacent to collagen fibrils (arrow heads) and in direct vicinity of phagocytes ($M\Phi$) or local

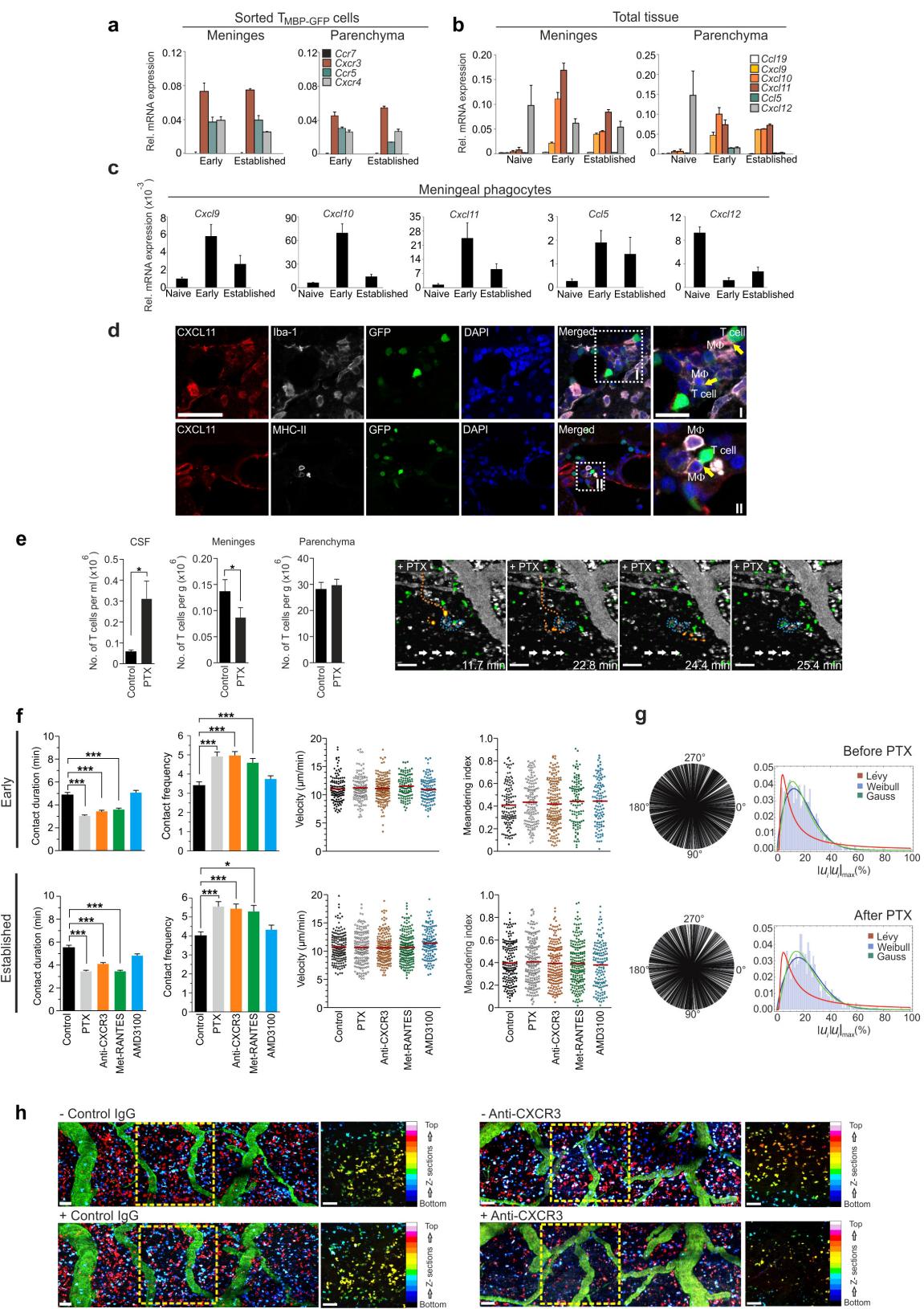
resident cells (M). Right, higher magnification of a lymphocyte in close contact with a macrophage. **c**, T_{MBP} cells do not follow chemotactic gradients towards meningeal phagocytes. Vector analysis of the migration steps of $T_{MBP-GFP}$ cells in relation to a fixed macrophage in areas with a radius (ρ) of 25, 50 or 100 μ m. The cosine of the angle between the macrophage–T-cell axes (AC) and the migration vector (AB) was calculated for each step. The values 1 or –1 correspond to a $T_{MBP-GFP}$ cell moving towards the meningeal phagocyte or in the opposite direction, respectively. Numbers indicate average direction \pm s.d. of the steps examined for each radius. The average is always close to zero and the standard deviation is very broad, excluding the presence of targeted migration. Representative data of 4 independent experiments for each radius including 4396, 15870 and 57535 steps for $\rho = 25$, $\rho = 50$ and $\rho = 100 \mu\text{m}$, respectively. **d**, Integrin ligands are highly expressed in meningeal phagocytes. Bar plots show quantitative PCR data for the indicated integrin ligands on meningeal phagocytes labelled with Texas Red sorted from naive animals or during the early or established phases of leptomeningeal T_{MBP} cell infiltration. House-keeping gene, β -actin (*Actb*). Mean \pm s.d. of duplicate measurements. Histogram plots show corresponding protein expression (flow cytometry) during the established phase of leptomeningeal T_{MBP} cell infiltration. Bottom, representative confocal images of leptomeningeal spots acquired in naive GFP^+ animals or in GFP^+ animals with established leptomeningeal T_{MBP} cell infiltration after *in vivo* staining with the indicated antibodies. Single fluorescent channels and merged pictures. The expression of the tested integrin ligands is enriched in the vicinity of the leptomeningeal blood vessels in naive tissue (left) but widely distributed during the established phase of leptomeningeal T_{MBP} cell infiltration (right).



Extended Data Figure 7 | See next page for figure caption.

Extended Data Figure 7 | Locomotion behaviour of T_{MBP} cells in the leptomeningeal milieu and its regulation by VLA-4/LFA-1 integrins. **a**, T_{MBP} cells express high level of integrins. VLA-4 (*Itga4*) and LFA-1 (*Itgal*) expression (quantitative PCR) on T_{MBP} -GFP cells from meninges or CNS parenchyma at the indicated time points of leptomeningeal T-cell infiltration. Representative data of 3 independent experiments. Data are mean \pm s.d. **b**, VLA-4 is expressed in the active conformation on T_{MBP} cells in the leptomeninges but not in the CSF. HUTS-4 antibody (directed against the activated conformation of $\beta 1$ -integrins), activation-independent anti-VLA-4 antibody or control antibody were injected i.t. into Lewis rats during the established phase of leptomeningeal T-cell infiltration. T_{MBP} -GFP cells were isolated from meninges and CSF 4 h later; protein expression was measured by flow cytometry. Percentages of positive T cells are indicated. Representative data of 3 independent experiments. **c**, Integrin blockade accelerates T-cell migration. Dot plots show T_{MBP} -GFP cell velocity before (control) and 4 h after i.t. treatment with the indicated anti-integrin monoclonal antibodies. Data from 30 min TPLSM recordings on leptomeninges during the early phase of T_{MBP} cell leptomeningeal infiltration. Data are mean values of individual T cells from at least 3 independent experiments including 674 cells (Kruskal–Wallis ANOVA followed by Dunn's multiple comparison). **d**, T_{MBP} cell velocity is not influenced by integrins in the absence of CSF flow. Velocities of T_{MBP} -GFP cells in the leptomeningeal milieu analysed during the established phase of leptomeningeal T_{MBP} cell infiltration after interruption of the CSF flow in the presence or absence (control) of anti-integrin monoclonal antibodies. The interruption of the CSF flow was achieved by removal of the arachnoidea causing the efflux of the CSF

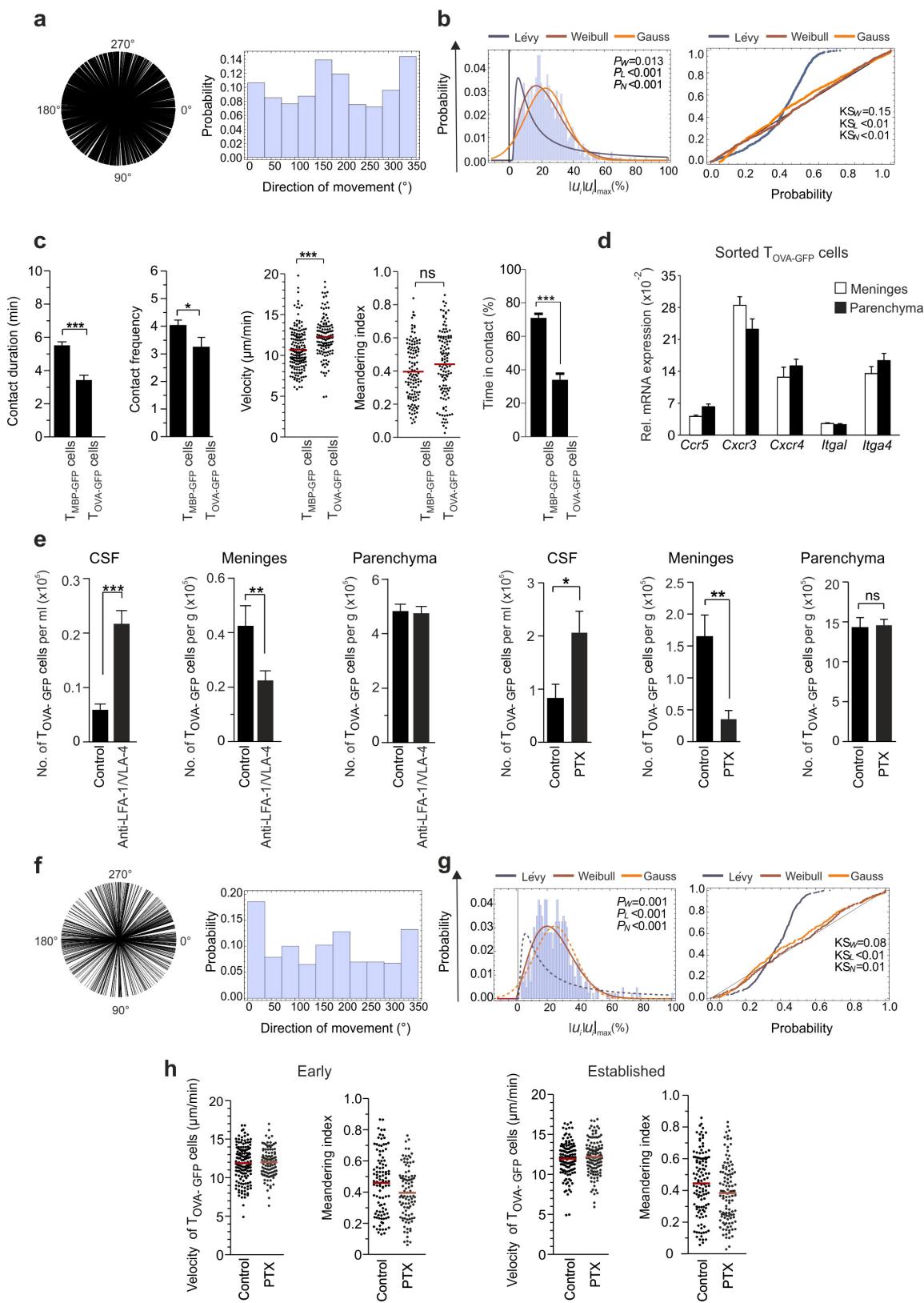
from the subarachnoidal space. 30 min TPLSM recordings. Data are mean values of individual T cells from 3 independent experiments including 166 cells (two-tailed Mann–Whitney *U*-test). **e**, Integrin blockade does not change the motility pattern of T_{MBP} cells. Mathematical analyses of at least 10 TPLSM recordings for each treatment (as in Fig. 3a, b). The Brownian walk of effector T_{MBP} -GFP cells was not changed after i.t. application of anti-LFA-1/VLA-4 monoclonal antibodies. **f**, Effects of integrins on T_{MBP} cell adhesion to the leptomeninges during the early phase of meningeal infiltration resemble those during the established phase. 30 min intravital TPLSM recordings on leptomeninges during the early phase of leptomeningeal T_{MBP} cell infiltration. Animals were treated i.t. with PBS (control) or with the indicated anti-integrin monoclonal antibodies. Bar plots show number of T_{MBP} -GFP cells 4 h after treatment in CSF and meninges (flow cytometry). At this time point no T_{MBP} -GFP cells were detectable in the CNS parenchyma (Kruskal–Wallis ANOVA followed by Dunn's multiple comparison test). Images show representative stills from 30 min TPLSM recordings and corresponding time projections of T-cell tracks before and 4 h after the indicated treatment. T-cell detachment is indicated by the reduction of T_{MBP} cell trajectories after i.t. monoclonal antibody treatment. Blue, T_{MBP} -Lifeact–Turquoise² cells. Scale bar, 50 μ m. Representative data of 2 independent experiments. **g**, Integrin blockade increases the turnover of T_{MBP} cells in the leptomeninges. Quantification of photoconverted T_{MBP} -Dendra₂ cells after i.t. application of anti-LFA-1/VLA-4 monoclonal antibodies or control IgG (control) during TPLSM of the established phase of leptomeningeal T_{MBP} -Dendra₂ cell infiltration. Representative kinetic of at least 3 independent experiments. **P* < 0.05, ***P* < 0.01, ****P* < 0.001 (c, f).



Extended Data Figure 8 | See next page for figure caption.

Extended Data Figure 8 | Role of chemokines in T_{MBP} cell motility in the leptomeningeal milieu. **a**, Chemokine receptor expression profile of T_{MBP} cells during EAE. Quantitative PCR data for the indicated chemokine receptors from $T_{MBP-GFP}$ cells isolated from spinal cord meninges and parenchyma during the early and established phases of leptomeningeal T_{MBP} cell infiltration. Representative data of 3 independent experiments. Data are mean \pm s.d. of replicate measurements. **b, c**, Inflammatory chemokines are upregulated in the CNS milieus and in meningeal macrophages during EAE. Quantitative PCR in naive animals or in animals during the early and established phases of leptomeningeal T_{MBP} cell infiltration. Chemokine ligand expression either on meninges and CNS parenchyma (**b**) or on Texas-Red-labelled meningeal phagocytes isolated by flow cytometry (**c**). House-keeping gene, β -actin (*Actb*) (**a–c**). Representative data of 3 different experiments. Data are mean \pm s.d. **d**, T_{MBP} cells establish direct contact with CXCL11⁺, MHC-II⁺ meningeal phagocytes. Leptomeninges during early leptomeningeal T_{MBP} cell infiltration stained for CXCL11, Iba-1 and DAPI (upper row) or CXCL11, MHC-II and DAPI (lower row). Green, $T_{MBP-GFP}$ cells. Scale bars, 50 μ m. In the magnifications (right) yellow arrows indicate direct contacts between $T_{MBP-GFP}$ cells and CXCL11⁺ meningeal phagocytes. Scale bars, 20 μ m. **e**, $G\alpha_i$ signalling blockade induces a release of T_{MBP} cells from the leptomeninges into the CSF. Left, plots show flow cytometry quantification of $T_{MBP-GFP}$ cells during established leptomeningeal T_{MBP} cell infiltration in the indicated compartments 4 h after i.t. treatment with PBS (control) or PTX. Right, intravital image of spinal cord leptomeninges recorded 4 h after i.t. injection of PTX during the established phase of leptomeningeal T_{MBP} cell infiltration. During the recording time, two $T_{MBP-GFP}$ cells highlighted in orange were released into the CSF. Grey (false colour), leptomeningeal blood vessels, phagocytes; green, $T_{MBP-GFP}$ cells; orange dotted lines, tracks of individual $T_{MBP-GFP}$ cells before being washed away into the CSF; blue dotted lines, contour of individual resident phagocytes; white arrows, direction of CSF flow. Scale bars, 50 μ m. **f**, Interference

with chemokine signalling affects T_{MBP} cell contacts with leptomeningeal phagocytes but not T_{MBP} cell velocity and meandering index. 30 min time-lapse recordings during the early (upper plots) and established phases (lower plots) of leptomeningeal T_{MBP} cell infiltration. Bar plots show mean contact duration and contact frequencies (early and established phases: 968 and 498 motile cells, respectively) between $T_{MBP-GFP}$ cells and Texas-Red-labelled meningeal phagocytes before (control) and 4 h after the indicated i.t. treatments. Data are mean \pm s.e.m. of 3 independent experiments. Dot plots show $T_{MBP-GFP}$ cell velocity and meandering index before (control) and 4 h after the indicated i.t. treatment. Data are mean values of individual T cells from at least 3 independent experiments per treatment and time point (early phase 1506 cells and established phase 860 cells; Kruskal–Wallis ANOVA followed by Dunn's multiple comparison test). **g**, $G\alpha_i$ blockade does not change the migratory pattern of T_{MBP} cells. Motility pattern of $T_{MBP-GFP}$ cells in the leptomeninges during the established phase of leptomeningeal T_{MBP} cell infiltration before and 4 h after PTX treatment (as in Fig. 3a,b). Directional analysis and probability plots of T cell displacement for the Lévy (red), Weibull (blue) and normal (Gauss, green) distributions. The $T_{MBP-GFP}$ cells maintained their Brownian walk after PTX treatment. Analysis of at least eight TPLSM recordings. **h**, CXCR3 blockade alone recapitulates the effect of $G\alpha_i$ inhibition on $T_{MBP-GFP}$ cells released from the meninges into the CSF. Representative TPLSM overviews of spinal cord leptomeninges before and 8 h after i.v. injection of isotype control antibodies (control IgG) or anti-CXCR3 monoclonal antibody during established leptomeningeal T_{MBP} cell infiltration. Magnification of individual regions originating from the overviews illustrate the position of the effector T cells in the z-axis in the leptomeningeal space via a colour-code scale. Note, $T_{MBP-GFP}$ cells are preferentially detached from the surface of the pia mater. Green, leptomeningeal blood vessels; red, phagocytes; blue, $T_{MBP-Lifact-Turquoise2}$ cells. Scale bars, 50 μ m. $*P < 0.05$, $***P < 0.001$ (**e, f**).

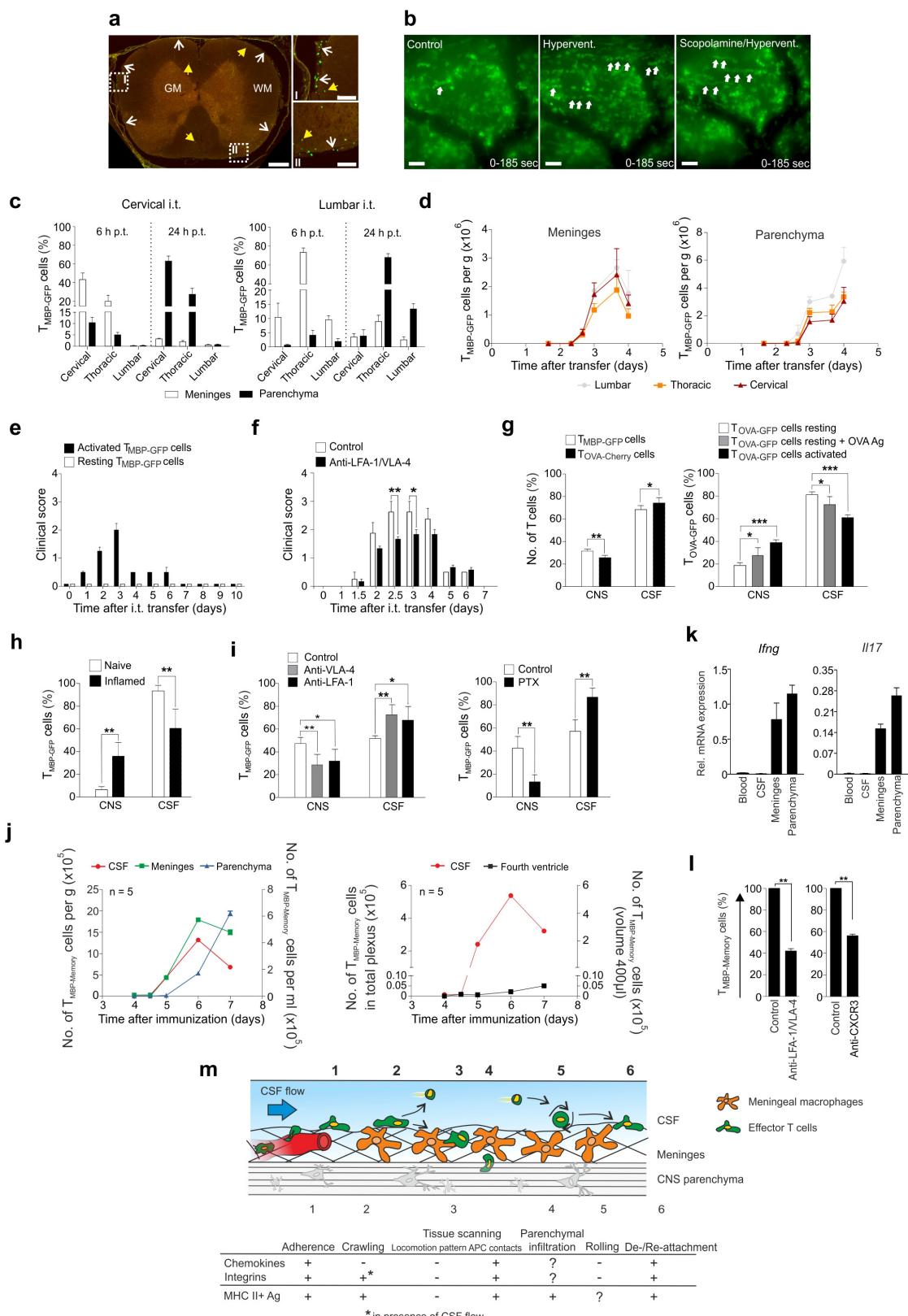


Extended Data Figure 9 | See next page for figure caption.

Extended Data Figure 9 | Motility pattern of brain-non-reactive Tova cells in the leptomeninges and their detachment into the CSF.

a, b, The motility pattern of T_{OVA} cells in the leptomeninges resembles that of T_{MBP} cells. $T_{OVA-GFP}$ cell motility in the leptomeninges during established inflammation (TPLSM 3.5 days after co-transfer with non-labelled T_{MBP} cells). **a**, T_{OVA} cells do not follow a preferential direction. Directional analyses of the $T_{OVA-GFP}$ cell tracks (as in Fig. 3a). Directions of in-plane movement (left) and associated probability for an angle varying within [0; 360] degrees in the fitting plane (right). **b**, T_{OVA} cells move in a Brownian random walk. Motility pattern of $T_{OVA-GFP}$ cells (as in Fig. 3b). Analysis of at least 6 TPLSM recordings. **c**, $T_{OVA-GFP}$ cells are less adhesive to meningeal structures in comparison to their myelin-reactive counterparts. Mean contact duration (417 contacts per 127 cells); contact frequencies (127 cells) between MBP- or OVA-reactive T cells and meningeal phagocytes; mean velocities (304 cells) and meandering indices (236 cells) of MBP- or OVA-reactive T cells; percentage of time spent by these T cells in contact with meningeal phagocytes during the 30 min recording. Results from intravital TPLSM recordings during established meningeal inflammation on spinal cord leptomeninges, after transfer of $T_{MBP-GFP}$ cells or of $T_{OVA-GFP}$ cells together with non-labelled T_{MBP} cells. Data are mean values \pm s.e.m. of 3 independent experiments per transfer (two-tailed Mann-Whitney *U*-test). **d**, Chemokine receptors and integrin expression in $T_{OVA-GFP}$ cells. $T_{OVA-GFP}$ cells transferred into Lewis rats together with non-labelled T_{MBP}

cells were sorted from meninges and CNS parenchyma during established meningeal inflammation. Quantitative PCR. House-keeping gene, β -actin (*Actb*). Representative data of 3 independent experiments. Data are mean \pm s.d. of duplicate measurements. **e**, Integrin or $G\alpha_i$ signalling interference induces a release of $T_{OVA-GFP}$ cells from the leptomeninges into the CSF. $T_{OVA-GFP}$ cells were co-injected with non-labelled T_{MBP} cells. During established meningeal inflammation (3.5 days after transfer) animals were treated i.t. with PBS (control), anti-LFA-1/VLA-4 monoclonal antibodies (left) or PTX (right). Absolute numbers of $T_{OVA-GFP}$ cells in CSF, spinal cord meninges and parenchyma 4 h after treatment. Data are mean \pm s.e.m. of representative data from at least 3 independent experiments per treatment including 3 animals per group (two-tailed Mann-Whitney *U*-test). **f-h**, $G\alpha_i$ signalling interference does not change the motility pattern of T_{OVA} cells. **f, g**, The Brownian walk of T_{OVA} cells was not changed after PTX treatment. Directional analyses and motility pattern of $T_{OVA-GFP}$ cells 4 h after i.t. injection of PTX. Mathematical analyses of at least 6 TPLSM recordings (as in Fig. 3a, b) 4 days after co-transfer with T_{MBP} cells. **h**, Interference with chemokine signalling does not affect T_{OVA} cell straightness and velocity. Mean velocities (498 cells) and meandering indices (428 cells) of $T_{OVA-GFP}$ cells at the depicted time points of leptomeningeal T-cell inflammation. 30 min TPLSM recordings were acquired before and 4 h after i.t. treatment with PTX. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$ (c-e).



Extended Data Figure 10 | See next page for figure caption.

Extended Data Figure 10 | Reattachment of T_{MBP} cells from the CSF to the leptomeninges and its regulation by integrins, chemokines and T-cell activation/CNS inflammation. **a**, Distribution of reattached T_{MBP} cells resembles that of early EAE lesions. T_{MBP-GFP} cells are located diffusely in the meninges and the adjacent spinal cord parenchyma after reattachment from the CSF. T_{MBP-GFP} cells retrieved from CNS tissue on day 3 after transfer were injected i.t. into naive animals. Fluorescence microscopy images of the fixed cervical spinal cord 20 h later. Insets, magnifications of areas of interest (right; I, II). Arrows, T_{MBP-GFP} cells in the leptomeninges (white) and CNS parenchyma (yellow). Green, T_{MBP-GFP} cells. Scale bars, 100 μ m. **b**, Increased T_{MBP} cell rolling and floating after forced ventilation indicate that respiration is a major driving force for the spinal cord CSF flow. Time-lapse video microscopy of T_{MBP-GFP} cells during the established phase of leptomeningeal T_{MBP} cell infiltration performed under 3 different conditions: (1) during standard conditions (control; respiration rate: 81 bpm; cardiac rate: 230 bpm); (2) during hyperventilation (hypervent; respiration rate: 100 bpm; cardiac rate: 200 bpm); and (3) during hyperventilation following administration of methylscopolamine (0.05 mg kg⁻¹) to block the hyperventilation-induced vagal influence on the heart (Scopolamine/hypervent; respiration rate: 100 bpm; cardiac rate: 230 bpm). Hyperventilation induced a strong increase in rolling/floating T_{MBP-GFP} cells that was not changed after methylscopolamine. 185 s time projections. Green, T_{MBP-GFP} cells; white arrows, representative T cells rolling/floating in the CSF. Scale bars, 50 μ m. Representative data of 3 independent experiments. **c**, T_{MBP} cell transport in the CSF after i.t. injection. Distribution of T_{MBP-GFP} cells in different levels of the spinal cord tissues after localized injection of the cells into the cisterna magna or the subarachnoidal space of the lumbar spinal cord. Flow cytometry analyses. Relative numbers of T_{MBP-GFP} cells 6 and 24 h after i.t. transfer in the indicated CNS compartments. Combined results of 2 independent experiments for each i.t. injected site ($n = 14$). **d**, T_{MBP} cell infiltration in the different levels of the spinal cord during transfer EAE. Quantification (flow cytometry) of T_{MBP-GFP} cells in the meninges or parenchyma of the indicated parts of the spinal cord. Data are mean \pm s.e.m. of representative data from 2 independent experiments per treatment including 2 or 3 animals per group ($n = 41$). **e**, Activated but not resting T_{MBP} cells induce EAE after i.t. transfer. T_{MBP-GFP} cells activated by antigenic stimulation or resting cells were i.t. transferred into naive animals. Clinical scores at the indicated time points. Representative results of 3 independent experiments with 4–6 animals per group. Data are mean \pm s.d. ($n = 28$). **f**, Interference with integrin binding reduces the encephalitogenic potential of i.t. transferred T_{MBP-GFP} cells. T_{MBP-GFP} cell blasts were pre-treated with either an IgG control antibody or a combination of anti-LFA-1/VLA-4 monoclonal antibodies before i.t. transfer into naive animals. Clinical scores. Representative results of 3 independent experiments with 2 or 3 animals per group. Data are mean \pm s.e.m. ($n = 16$, two-tailed Mann–Whitney *U*-test). **g**, T_{MBP} cells invade the inflamed CNS tissue more efficiently than T_{OVA} cells. Left, the same number of non-activated T_{MBP-GFP} and T_{OVA-Cherry} cells were co-injected i.t. into animals at the onset of clinical EAE, that is, 3 days after i.v. transfer with unlabelled T_{MBP-GFP} cells. 20 h after i.t. injection, the entry of T_{MBP-GFP} and T_{OVA-Cherry} cells into the CNS tissue

(including spinal cord meninges and parenchyma) was quantified by flow cytometry. Representative results of 2 independent experiments with 4 animals per group. Data are mean \pm s.d. ($n = 8$). Right, resting or *in vitro* activated T_{OVA-GFP} cells were i.t. injected into naive animals. In addition, a group of animals receiving resting T_{OVA-GFP} cells were co-injected intrathecally with 25 μ g of OVA antigen. 20 h after i.t. injection, T_{OVA-GFP} cells in the spinal cord tissue were quantified. Representative of 2 independent experiments with 2 animals per group. Data are mean \pm s.d. ($n = 12$, two-tailed Mann–Whitney *U*-test (left) or Kruskal–Wallis ANOVA followed by Dunn's multiple comparison test (right)).

h, Inflammation of the CNS tissue increases T_{MBP} cell entry from the CSF. CSF-derived T_{MBP-GFP} cells were injected i.t. either into naive animals or into animals at the onset of clinical EAE, that is, 3 days after i.v. transfer of non-labelled T_{MBP} cells (that is, inflamed condition). Cell quantifications of T_{MBP-GFP} cells in the CNS tissue and CSF by flow cytometry 20 h later when the re-transferred T cells had maximally infiltrated the inflamed CNS of the recipient animals. Data are mean \pm s.d. of representative results of 3 independent experiments ($n = 12$, two-tailed Mann–Whitney *U*-test). **i**, Blocking of integrin or G α_i -signalling reduces T_{MBP} cell migration from the CSF into the inflamed CNS tissue. CNS-derived T_{MBP-GFP} cells were pre-treated with anti-VLA-4 or anti-LFA-1 monoclonal antibodies or with isotype control IgG antibody (control) (left) or PTX or PBS as control (right). Cells were then injected i.t. in animals at the onset of EAE and quantified as in **h**. Data are mean \pm s.d. of representative results of 3 independent experiments ($n = 18$, Kruskal–Wallis ANOVA followed by Dunn's multiple comparison test (left) or two-tailed Mann–Whitney *U*-test, right). **j–l**, In active EAE induced by reactivated memory T_{MBP} cells, trafficking of the memory T_{MBP} cells between the distinct CNS compartments follows the same rules as the trafficking of T_{MBP} cells during transfer EAE. **j**, Memory T_{MBP} cells accumulate simultaneously in the meninges and the CSF before they occur in the choroid plexus. Quantification (flow cytometry) of GFP⁺ T_{MBP-Memory} cells in CSF, spinal cord meninges and parenchyma (left), or in CSF and choroid plexus of the fourth ventricle (right) at the indicated time points after immunization of 10-week-old memory animals. Data are mean \pm s.d. of 3 independent experiments ($n = 13$). **k**, CSF-derived memory T_{MBP} cells present a low activation profile. IFN γ and IL-17 expression (quantitative PCR) in GFP⁺ memory T_{MBP} cells 5 days after immunization. House-keeping gene, β -actin (*Actb*). Representative data of 3 independent experiments. Data are mean \pm s.d. of duplicate measurements. **l**, Integrin and chemokine blockade reduces the number of T_{MBP-Memory} cells in the leptomeninges during active EAE. Intravital TPLSM recordings of spinal cord leptomeninges from 10-week-old memory animals 5 days after immunization. Quantification of GFP⁺ memory T_{MBP} cells in the acquired images before (control) and 4 h after i.t. treatment of either anti-LFA-1/VLA-4 or anti-CXCR3 blocking monoclonal antibodies. Data are mean values \pm s.e.m. from 3 independent experiments including 5,947 cells (two-tailed Mann–Whitney *U*-test). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ (**g–l**). **m**, Schematic representation of the migratory behaviour of effector T cells in the leptomeningeal milieu and the roles of chemokines, integrins and activation in controlling the T-cell migration steps. Question marks indicate unresolved points.

NEK7 is an essential mediator of NLRP3 activation downstream of potassium efflux

Yuan He¹, Melody Y. Zeng¹, Dahai Yang^{1,2}, Benny Motro³ & Gabriel Núñez¹

Inflammasomes are intracellular protein complexes that drive the activation of inflammatory caspases¹. So far, four inflammasomes involving NLRP1, NLRP3, NLRC4 and AIM2 have been described that recruit the common adaptor protein ASC to activate caspase-1, leading to the secretion of mature IL-1 β and IL-18 proteins^{2,3}. The NLRP3 inflammasome has been implicated in the pathogenesis of several acquired inflammatory diseases^{4,5} as well as cryopyrin-associated periodic fever syndromes (CAPS) caused by inherited NLRP3 mutations^{6,7}. Potassium efflux is a common step that is essential for NLRP3 inflammasome activation induced by many stimuli^{8,9}. Despite extensive investigation, the molecular mechanism leading to NLRP3 activation in response to potassium efflux remains unknown. Here we report the identification of NEK7, a member of the family of mammalian NIMA-related kinases (NEK proteins)¹⁰, as an NLRP3-binding protein that acts downstream of potassium efflux to regulate NLRP3 oligomerization and activation. In the absence of NEK7, caspase-1 activation and IL-1 β release were abrogated in response to signals that activate NLRP3, but not NLRC4 or AIM2 inflammasomes. NLRP3-activating stimuli promoted the NLRP3–NEK7 interaction in a process that was dependent on potassium efflux. NLRP3 associated with the catalytic domain of NEK7, but the catalytic activity of NEK7 was shown to be dispensable for activation of the NLRP3 inflammasome. Activated macrophages formed a high-molecular-mass NLRP3–NEK7 complex, which, along with ASC oligomerization and ASC speck formation, was abrogated in the absence of NEK7. NEK7 was required for macrophages containing the CAPS-associated NLRP3(R258W) activating mutation to activate caspase-1. Mouse chimaeras reconstituted with wild-type, *Nek7*^{-/-} or *Nlrp3*^{-/-} haematopoietic cells showed that NEK7 was required for NLRP3 inflammasome activation *in vivo*. These studies demonstrate that NEK7 is an essential protein that acts downstream of potassium efflux to mediate NLRP3 inflammasome assembly and activation.

To understand the signalling mechanism of NLRP3 inflammasome activation, we sought to identify proteins that interact with NLRP3 after inflammasome activation. To purify NLRP3 protein complexes, we generated a triple-tagged NLRP3 (NLRP3-SFP) fused with three tags in the carboxyl terminus: S-tag, Flag (for detection), and a streptavidin-binding tag. Reconstitution of *Nlrp3*^{-/-} immortalized bone-marrow-derived macrophages (iBMDMs) with NLRP3-SFP restored ATP-induced caspase-1 activation and IL-1 β release (Extended Data Fig. 1a). We treated lipopolysaccharide (LPS)-primed reconstituted iBMDMs with ATP to induce NLRP3 activation, and searched for interacting partners of NLRP3 using liquid chromatography–mass spectrometry. The analysis revealed NEK7 as a major interacting partner of NLRP3 (Fig. 1a). NLRP3 did not associate with NEK6, a NEK7-related parologue, or NEK9, another member of the NEK family¹⁰ (Extended Data Fig. 1b). The NLRP3–NEK7 interaction was confirmed by pull-down assays using streptavidin beads or immunoprecipitation (Fig. 1b, c and Extended Data Fig. 1b–d).

Notably, the NLRP3–NEK7 interaction was slightly increased by LPS priming, but was clearly enhanced after ATP stimulation (Fig. 1a–c and Extended Data Fig. 1b–d). The interaction of NLRP3 with NEK7 was independent of ASC, caspase-1 or caspase-11 (Extended Data Fig. 1d). To determine the regions within NLRP3 that associate with

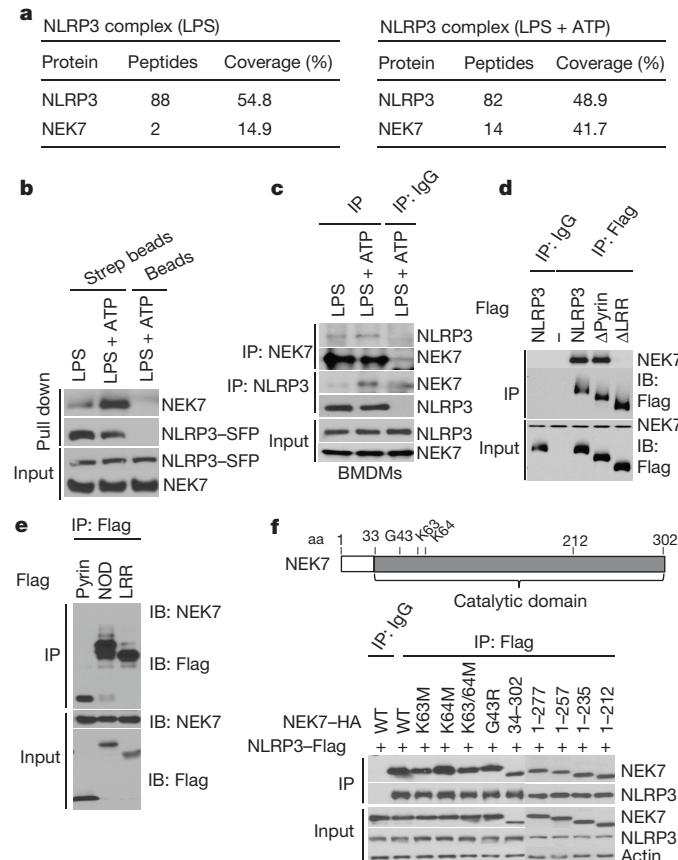


Figure 1 | NEK7 interacts with NLRP3. **a**, Mass spectrometry analysis of NLRP3 and NEK7 peptides after purification of NLRP3-associated proteins. **b**, NLRP3-SFP was pulled down and immunoblotted with indicated antibodies. Strep, streptavidin. **c**, LPS-primed BMDMs were left unstimulated or stimulated with ATP for 30 min. Cell lysates were immunoprecipitated (IP) and immunoblotted (IB) with indicated antibodies. **d, e**, Wild-type or mutant NLRP3 (Δ Pyrin or Δ LRR (leucine-rich repeats)) was expressed in HEK293T cells, immunoprecipitated and analysed by immunoblotting. **f**, Wild-type or mutant NEK7 was co-expressed with Flag-tagged NLRP3 in HEK293T cells, immunoprecipitated and analysed by immunoblotting. Whole-cell lysates are shown as the input. aa, amino acids; HA, haemagglutinin. Results are representative of at least three independent experiments. See Supplementary Fig. 1 for gel source data.

¹Department of Pathology and Comprehensive Cancer Center, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. ²The State Key Laboratory of Bioreactor Engineering, East China University of Science and Technology, Shanghai 200237, China. ³The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat-Gan 52900, Israel.

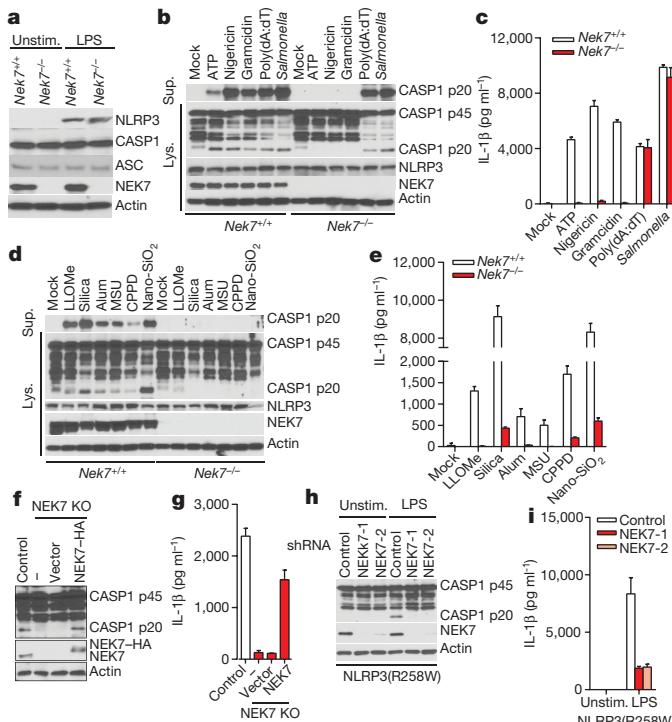


Figure 2 | NEK7 deficiency specifically abrogates the activation of the NLRP3 inflammasome. **a**, BMDMs were left untreated or stimulated with LPS, and cell lysates were immunoblotted with indicated antibodies. **b, d**, Caspase-1 (CASP1) in the supernatant (sup.) and cell lysate (lys.) was analysed in stimulated *Nek7*^{+/+} and *Nek7*^{-/-} macrophages treated with ATP, nigericin, gramicidin, poly(dA:dT) or *Salmonella* (**b**) or with LLOMe, silica, Alum, MSU, calcium pyrophosphate dihydrate (CPPD) or nano-SiO₂ (**d**). Mock represents macrophages primed with LPS without further stimulation. **c, e**, IL-1 β release was measured. **f, g**, CRISPR-Cas9-generated NEK7-deficient (NEK7 knockout (KO)) iBMDMs were transduced with control or pHIV-NEK7 lentivirus, and stimulated with LPS plus nigericin. Caspase-1 activation (**f**) and IL-1 β release (**g**) were analysed. **h, i**, Macrophages were transduced with control lentivirus or shRNA lentiviruses targeting NEK7, and stimulated with LPS. Caspase-1 activation (**h**) and IL-1 β release (**i**) were analysed. IL-1 β release data (**c, e, g, i**) are expressed as mean values. Error bars denote s.d. of triplicate wells. Results are representative of three independent experiments. See Supplementary Fig. 1 for gel source data.

NEK7, we expressed Flag-tagged wild-type or mutant NLRP3 in HEK293T cells. NEK7 interacted with wild-type or mutant NLRP3 lacking the amino-terminal pyrin domain, but not with NLRP3 lacking the carboxy-terminal leucine-rich repeats (Fig. 1d). Furthermore, NEK7 did not associate with the singly expressed pyrin domain, the centrally located nucleotide-binding domain (NOD) or leucine-rich repeats (Fig. 1e), indicating that both the NOD and leucine-rich repeats are involved in the interaction with NEK7. Conversely, the catalytic domain of NEK7, but not the N-terminal 33-amino-acid extension domain, interacted with NLRP3 (Fig. 1f). Further analysis showed that the N-terminal region (amino acid residues 34–212), but not the C-terminal region (residues 213–302), of the NEK7 catalytic domain¹¹ mediates the interaction with NLRP3 (Fig. 1f). However, mutations of the catalytic Lys63 and Lys64 or the Gly43 amino acid residue that abolish the catalytic activity of NEK7 (refs 12, 13) did not impair the NLRP3–NEK7 interaction (Fig. 1f). These results indicate that NEK7 interacts with NLRP3 and the interaction is enhanced in response to NLRP3-activating stimuli.

We next evaluated the requirement for NEK7 in NLRP3 inflammasome activation. Because NEK7 deficiency leads to either embryonic lethality or death of pups soon after birth¹⁴, we generated mouse chimaeras after transplanting fetal liver cells from *Nek7*^{+/+} or *Nek7*^{-/-} embryos into lethally irradiated recipient mice. BMDMs from mice reconstituted with *Nek7*^{-/-} cells lacked detectable expression of

NEK7, but expressed normal amounts of NLRP3, caspase-1 and ASC (Fig. 2a). Importantly, activation of caspase-1 and IL-1 β release induced by ATP, nigericin and toxin gramicidin, three stimuli that activate NLRP3, were abolished in *Nek7*^{-/-} BMDMs (Fig. 2b, c). By contrast, activation of caspase-1 and IL-1 β release in response to poly(deoxyadenylic-deoxythymidylic) acid (poly(dA:dT)) that activates the AIM2 inflammasome, or *Salmonella enterica* serovar Typhimurium (*Salmonella*) that activates the NLRC4 inflammasome, were not affected in *Nek7*^{-/-} BMDMs (Fig. 2b, c). Likewise, caspase-1 activation and IL-1 β release induced by particulate matter and the lysosome membrane damaging agent Leu-Leu-OME (LLOMe), were impaired in *Nek7*^{-/-} BMDMs (Fig. 2d, e). By contrast, TNF- α release induced by all tested stimuli was unaffected in *Nek7*^{-/-} BMDMs (Extended Data Fig. 2a, b). In addition, NLRP3-dependent caspase-1 activation and IL-1 β release induced by cytosolic LPS stimulation that activates the non-canonical inflammasome via caspase-11 also required NEK7 (Extended Data Fig. 2c, d). Consistent with previous studies^{15–17}, cytotoxicity induced by cytosolic LPS required caspase 11, but not NLRP3 or NEK7 (Extended Data Fig. 2e). To ensure that impaired NLRP3 activation in *Nek7*^{-/-} BMDMs was not secondary to abnormal mouse development, we deleted *Nek7* using CRISPR-Cas9 genome editing in iBMDMs. NLRP3 inflammasome activation induced by nigericin was abrogated in NEK7-deficient macrophages (Fig. 2f, g and Extended Data Fig. 3a–c). Importantly, re-expression of NEK7 in NEK7-deficient macrophages restored NLRP3 inflammasome activation (Fig. 2f, g). Similarly, knockdown of NEK7 by short hairpin RNAs (shRNAs) targeting NEK7 impaired caspase-1 activation and IL-1 β release, but not TNF- α production, in response to ATP, nigericin or silica (Extended Data Fig. 4a–d). We also depleted NEK7 in BMDMs containing the activating NLRP3(R258W) mutation corresponding to the human NLRP3(R260W) mutation that causes Muckle–Wells syndrome¹⁸. In agreement with previous studies¹⁹, treatment of NLRP3(R258W) BMDMs with LPS alone was sufficient to activate caspase-1 and IL-1 β release (Fig. 2h, i). Notably, caspase-1 activation and IL-1 β release elicited by LPS in NLRP3(R258W) BMDMs were impaired by NEK7 knockdown (Fig. 2h, i). These results indicate that NEK7 acts on or just downstream of both wild-type and CAPS-associated NLRP3 to regulate the inflammasome.

Stimulation of *Nek7*^{+/+} BMDMs with the NLRP3 activators ATP and nigericin, as well as poly(dA:dT) or *Salmonella*, induced rapid formation of large intracellular ASC aggregates called ASC specks in the cytosol (Fig. 3a, b). The formation of ASC specks induced by ATP or nigericin was abrogated, but unperturbed when induced by poly(dA:dT) or *Salmonella*, in *Nek7*^{-/-} BMDMs (Fig. 3a, b). Consistently, ASC oligomerization triggered by stimulation with several NLRP3 activators, but not poly(dA:dT) or *Salmonella* infection, was abolished in *Nek7*^{-/-} BMDMs or greatly reduced in BMDMs with knockdown of NEK7 (Fig. 3c and Extended Data Fig. 5a–c). Activated inflammasomes assemble into high-molecular-mass multiprotein complexes^{1,20}. To assess NLRP3 inflammasome assembly, wild-type and *Nek7*^{-/-} BMDMs were stimulated with nigericin or ATP, and digitonin-solubilized cell lysates were resolved by blue native polyacrylamide gel electrophoresis (PAGE), and the blots were then immunoblotted with anti-NLRP3 and anti-NEK7 antibodies. A large oligomeric complex (>1,000 kilodaltons (kDa)) containing NLRP3 and NEK7 was induced in wild-type BMDMs after stimulation with ATP or nigericin, and this was greatly reduced or absent in stimulated *Nek7*^{-/-} and *Nlrp3*^{-/-} cells (Fig. 3d). To resolve the formation of NLRP3 oligomers better, we separated the samples in the first dimension by blue native PAGE and then in a second dimension by SDS-PAGE. Immunoblotting revealed that NEK7 was indeed present in a high-molecular-mass NLRP3 complex induced by ATP or nigericin in primary wild-type BMDMs, which was absent in unstimulated wild-type cells and stimulated *Nek7*^{-/-} or *Nlrp3*^{-/-} cells (Fig. 3e). To determine whether the kinase activity of NEK7 is important for the regulation of the NLRP3 inflammasome, we expressed wild-type

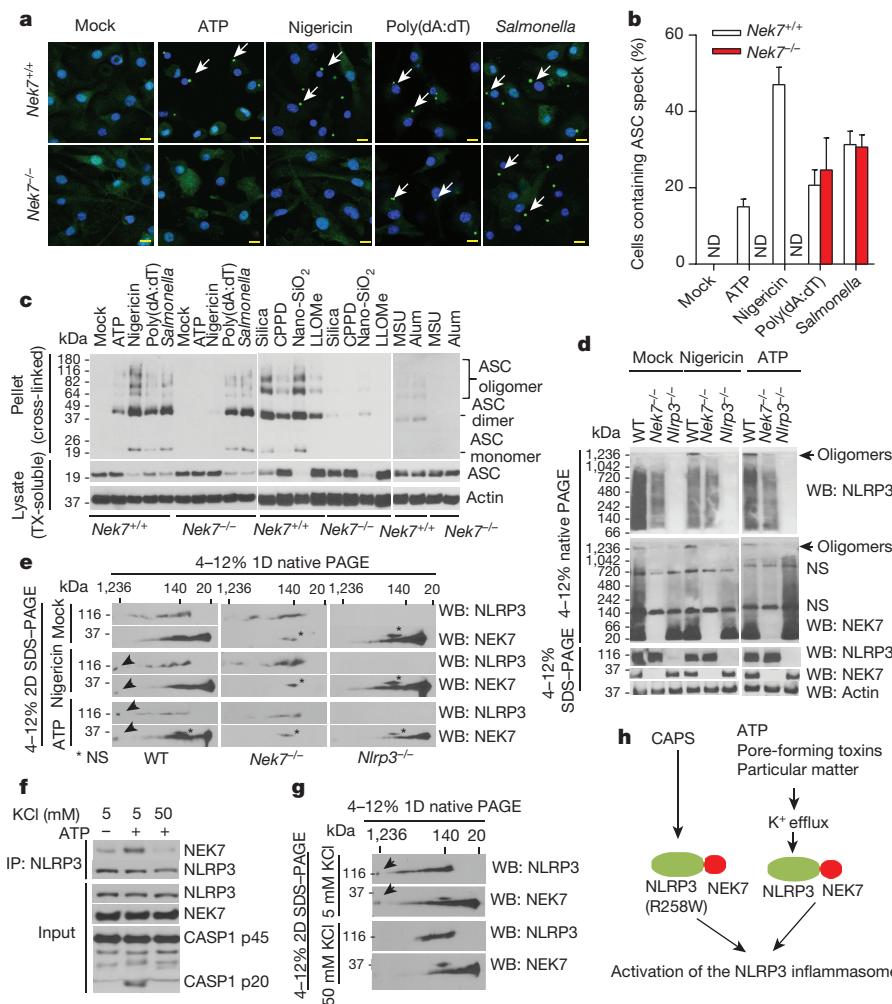


Figure 3 | NEK7 is required for NLRP3 oligomerization and ASC speck formation downstream of potassium efflux. **a, b**, Representative immunofluorescence images and quantification of endogenous ASC specks (arrows). ND, not detected. Data show representative results from three combined independent experiments. Scale bars, 10 μ m. Error bars indicate s.d. **c**, ASC oligomerization induced by indicated stimuli in *Nek7*^{+/+} and *Nek7*^{-/-} LPS-primed macrophages. Mock denotes stimulation with PBS. TX, Triton X-100. **d**, Indicated LPS-primed macrophages were stimulated with PBS (mock), ATP or nigericin. Analyses by blue native PAGE or SDS-PAGE, and immunoblotting. WB, western blot. **e**, Cell lysates were separated by a first dimension of blue native PAGE followed by a second dimension of SDS-PAGE. NS, nonspecific band. **f, g**, NEK7–NLRP3 interactions and complex formation in the presence of 5 mM or 50 mM extracellular KCl. **h**, Proposed model for mechanism of NEK7-mediated NLRP3 inflammasome activation. Results are representative of at least three independent experiments. See Supplementary Fig. 2 for gel source data.

or mutant NEK7 with mutations at the critical catalytic lysine residues Lys63 and Lys64 or glycine residue Gly43 in *Nek7*^{-/-} iBMDMs by lentivirus-mediated transduction. All of these NEK7 mutations have been found to abolish the kinase activity of NEK7 (refs 12, 13). Importantly, the ability of nigericin to induce caspase-1 activation and IL-1 β release was not impaired in *Nek7*^{-/-} iBMDMs reconstituted with these NEK7 mutant proteins (Extended Data Fig. 6a–c). In addition, shRNA depletion of NEK9, the kinase that phosphorylates and activates NEK7 (refs 11, 21), or of NEK7-related kinase NEK6 did not impair the activation of the NLRP3 inflammasome (Extended Data Fig. 7a–e). NEK7 regulates microtubule dynamics^{22,23}. However, treatment with nocodazole or colchicine, two inhibitors of tubulin polymerization, did not inhibit NLRP3 inflammasome activation and had a marginal inhibitory effect on IL-1 β and TNF- α release (Extended Data Fig. 8a–e). Furthermore, inhibition of tubulin polymerization did not affect the NLRP3–NEK7 interaction (Extended Data Fig. 8f). Inhibition of Bruton's tyrosine kinase (BTK) that has been implicated in NLRP3 inflammasome activation²⁴ did not disrupt the interaction between NEK7 and NLRP3 (Extended Data Fig. 9a, b). Additionally, we could not detect BTK in NLRP3 complexes before or after ATP stimulation (Extended Data Fig. 9b). Stimulation of *Nek7*^{+/+} and *Nek7*^{-/-} BMDMs with ATP, nigericin or gramicidin induced comparable levels of potassium efflux (Extended Data Fig. 10). Notably, the association of NLRP3 with NEK7 as well as caspase-1 activation triggered by ATP were blocked in the presence of 50 mM KCl that inhibits potassium efflux and NLRP3 activation (Fig. 3f). Importantly, the induced high-molecular-mass NLRP3 complex containing NEK7 was not seen in BMDMs cultured in 50 mM KCl (Fig. 3g). These results indicate that potassium efflux promotes the

NLRP3–NEK7 interaction that is required for NLRP3 inflammasome assembly and caspase-1 activation (Fig. 3h).

We assessed the role of NEK7 in the regulation of the inflammasome *in vivo*. We generated mouse chimaeras after transplanting fetal liver cells from wild-type, *Nek7*^{-/-} or *Nlrp3*^{-/-} embryos into lethally irradiated recipient mice, and IL-1 β production induced by LPS challenge was assessed in the serum^{25,26}. Administration of LPS to chimaeric mice reconstituted with wild-type fetal liver cells induced IL-1 β production, which was diminished similarly in chimaeric mice reconstituted with *Nek7*^{-/-} or *Nlrp3*^{-/-} haematopoietic cells (Fig. 4a). Consistent with additional NLRP3 function in radioresistant recipient cells²⁷, chimaeric *Nlrp3*^{-/-} recipients transplanted with *Nek7*^{-/-} or *Nlrp3*^{-/-} fetal liver cells showed further reductions in IL-1 β production (Fig. 4a). Notably, production of IL-6 and TNF- α in the sera of all chimaeric mice was comparable (Fig. 4b, c). Likewise, production of IL-1 β induced by intraperitoneal administration of monosodium urate (MSU) crystals, which is also largely mediated by the NLRP3 inflammasome²⁸, was reduced at comparable levels in chimaeric mice reconstituted with *Nek7*^{-/-} or *Nlrp3*^{-/-} fetal liver cells (Fig. 4d). These studies indicate that NEK7 is required for activation of the NLRP3 inflammasome *in vivo*.

We have shown that NEK7 is an essential factor that specifically and non-redundantly functions downstream of potassium efflux to regulate the activation of the NLRP3 inflammasome. Our results are in agreement with recent studies that identified NEK7 as a critical regulator of the NLRP3 inflammasome^{29,30}. Our studies suggest a model in which potassium efflux induced by NLRP3-activating stimuli triggers the association of NLRP3 with NEK7, leading to the assembly and activation of the NLRP3 inflammasome (Fig. 3h). Macrophages containing the CAPS-associated NLRP3(R258W) activating mutation that

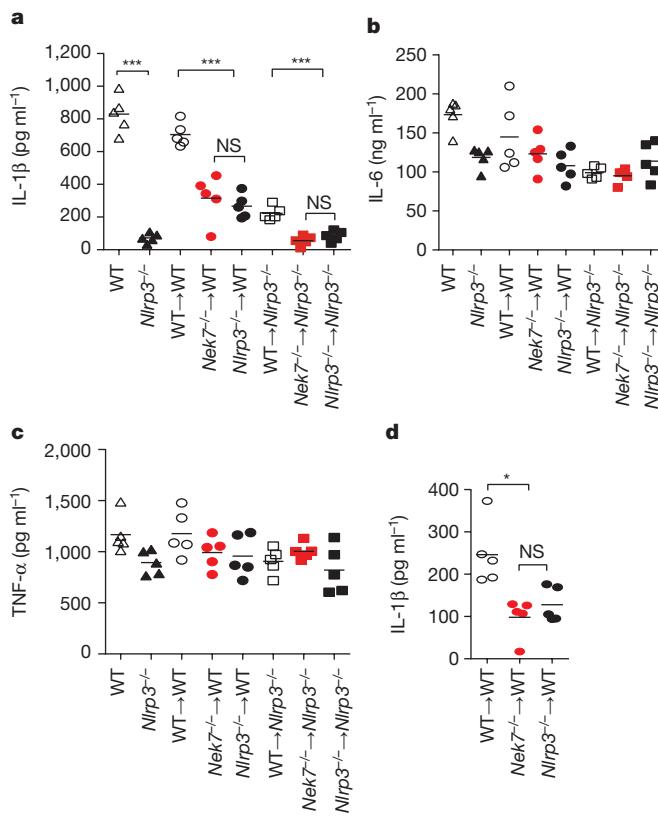


Figure 4 | NEK7 is required for activation of the NLRP3 inflammasome in vivo. **a–c**, Mouse serum cytokines IL-1 β (a), IL-6 (b) and TNF- α (c) were analysed after intraperitoneal injection of LPS. **d**, IL-1 β was analysed in peritoneal lavage fluids after intraperitoneal injection of MSU. Each symbol represents one mouse. Mean values are indicated by a horizontal bar. In the absence of LPS or MSU stimulation, the amounts of IL-1 β were undetectable. Results are representative of two independent experiments. NS, not significant. *P < 0.05; **P < 0.01; ***P < 0.001 (Kruskal-Wallis test).

does not require potassium efflux for inflammasome activation¹⁹ also required NEK7 for caspase-1 activation, suggesting that this mutant NLRP3 may be competent for NEK7 association in the absence of potassium efflux. NEK7 regulates microtubule dynamic instability and spindle assembly which required the NEK7 catalytic activity^{12,13,22,23}. By contrast, the catalytic activity of NEK7 is not required for NLRP3 activation. Further work is needed to understand the dual functions of NEK7. Taken together, our studies suggest that NEK7 could be a potential target of therapeutics to treat inflammatory diseases linked to NLRP3 inflammasome activation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 November 2015; accepted 5 January 2016.

Published online 27 January 2016.

- Martinon, F., Burns, K. & Tschopp, J. The inflammasome: A molecular platform triggering activation of inflammatory caspases and processing of proIL-1 β . *Mol. Cell* **10**, 417–426 (2002).
- Rathinam, V. A., Vanaja, S. K. & Fitzgerald, K. A. Regulation of inflammasome signaling. *Nature Immunol.* **13**, 333–342 (2012).
- Guo, H., Callaway, J. B. & Ting, J. P. Inflammasomes: mechanism of action, role in disease, and therapeutics. *Nature Med.* **21**, 677–687 (2015).
- Martinon, F., Mayor, A. & Tschopp, J. The inflammasomes: guardians of the body. *Annu. Rev. Immunol.* **27**, 229–265 (2009).
- Davis, B. K., Wen, H. & Ting, J. P. The inflammasome NLRs in immunity, inflammation, and associated diseases. *Annu. Rev. Immunol.* **29**, 707–735 (2011).
- Hoffman, H. M., Mueller, J. L., Broide, D. H., Wanderer, A. A. & Kolodner, R. D. Mutation of a new gene encoding a putative pyrin-like protein causes familial cold autoinflammatory syndrome and Muckle–Wells syndrome. *Nature Genet.* **29**, 301–305 (2001).

- Feldmann, J. et al. Chronic infantile neurological cutaneous and articular syndrome is caused by mutations in *CIAS1*, a gene highly expressed in polymorphonuclear cells and chondrocytes. *Am. J. Hum. Genet.* **71**, 198–203 (2002).
- Muñoz-Planillo, R. et al. K⁺ efflux is the common trigger of NLRP3 inflammasome activation by bacterial toxins and particulate matter. *Immunity* **38**, 1142–1153 (2013).
- Pétrilli, V. et al. Activation of the NALP3 inflammasome is triggered by low intracellular potassium concentration. *Cell Death Differ.* **14**, 1583–1589 (2007).
- Fry, A. M., O'Regan, L., Sabir, S. R. & Bayliss, R. Cell cycle regulation by the NEK family of protein kinases. *J. Cell Sci.* **125**, 4423–4433 (2012).
- Richards, M. W. et al. An autoinhibitory tyrosine motif in the cell-cycle-regulated Nek7 kinase is released through binding of Nek9. *Mol. Cell* **36**, 560–570 (2009).
- Yissachar, N., Salem, H., Tennenbaum, T. & Motro, B. Nek7 kinase is enriched at the centrosome, and is required for proper spindle assembly and mitotic progression. *FEBS Lett.* **580**, 6489–6495 (2006).
- O'Regan, L. & Fry, A. M. The Nek6 and Nek7 protein kinases are required for robust mitotic spindle formation and cytokinesis. *Mol. Cell. Biol.* **29**, 3975–3990 (2009).
- Salem, H. et al. Nek7 kinase targeting leads to early mortality, cytokinesis disturbance and polyploidy. *Oncogene* **29**, 4046–4057 (2010).
- Kayagaki, N. et al. Non-canonical inflammasome activation targets caspase-11. *Nature* **479**, 117–121 (2011).
- Kayagaki, N. et al. Noncanonical inflammasome activation by intracellular LPS independent of TLR4. *Science* **341**, 1246–1249 (2013).
- Hagar, J. A., Powell, D. A., Aachoui, Y., Ernst, R. K. & Miao, E. A. Cytoplasmic LPS activates caspase-11: implications in TLR4-independent endotoxic shock. *Science* **341**, 1250–1253 (2013).
- Neven, B. et al. Molecular basis of the spectral expression of *CIAS1* mutations associated with phagocytic cell-mediated autoinflammatory disorders CINCA/NOMID, MWS, and FCU. *Blood* **103**, 2809–2815 (2004).
- Meng, G., Zhang, F., Fuss, I., Kitani, A. & Strober, W. A mutation in the *Nlrp3* gene causing inflammasome hyperactivation potentiates Th17 cell-dominant immune responses. *Immunity* **30**, 860–874 (2009).
- Kofoed, E. M. & Vance, R. E. Innate immune recognition of bacterial ligands by NLRPs determines inflammasome specificity. *Nature* **477**, 592–595 (2011).
- Belham, C. et al. A mitotic cascade of NIMA family kinases. Nerk1/Nek9 activates the Nek6 and Nek7 kinases. *J. Biol. Chem.* **278**, 34897–34909 (2003).
- Cohen, S., Aizer, A., Shav-Tal, Y., Yanai, A. & Motro, B. Nek7 kinase accelerates microtubule dynamic instability. *Biochim. Biophys. Acta* **1833**, 1104–1113 (2013).
- Kim, S., Lee, K. & Rhee, K. NEK7 is a centrosomal kinase critical for microtubule nucleation. *Biochem. Biophys. Res. Commun.* **360**, 56–62 (2007).
- Ito, M. et al. Bruton's tyrosine kinase is essential for NLRP3 inflammasome activation and contributes to ischaemic brain injury. *Nature Commun.* **6**, 7360 (2015).
- Sutterwala, F. S. et al. Critical role for NALP3/CIAS1/cryopyrin in innate and adaptive immunity through its regulation of caspase-1. *Immunity* **24**, 317–327 (2006).
- Mariathasan, S. et al. Cryopyrin activates the inflammasome in response to toxins and ATP. *Nature* **440**, 228–232 (2006).
- Zaki, M. H. et al. The NLRP3 inflammasome protects against loss of epithelial integrity and mortality during experimental colitis. *Immunity* **32**, 379–391 (2010).
- Zhou, R., Tardivel, A., Thorens, B., Choi, I. & Tschopp, J. Thioredoxin-interacting protein links oxidative stress to inflammasome activation. *Nature Immunol.* **11**, 136–140 (2010).
- Schmid-Burgk, J. L. et al. A genome-wide CRISPR screen identifies NEK7 as an essential component of NLRP3 inflammasome activation. *J. Biol. Chem.* **291**, 103–109 (2015).
- Shi, H. et al. NLRP3 activation and mitosis are mutually exclusive events coordinated by NEK7, a new inflammasome component. *Nature Immunol.* <http://dx.doi.org/10.1038/ni.3333> (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank R. Muñoz-Planillo for potassium efflux assays, S. Varadarajan for plasmid construction, L. Burmeister for animal husbandry, J. Whitfield for ELISA assays, and V. Basrur for mass spectrometry analysis. Y.H. was supported by NIH training grant T32HL007517. M.Y.Z. was supported by NIH training grants T32HL007517 and T32DK049477. D.Y. was supported by the State Scholarship Fund from China Scholarship Council (no. 201306740018). This work was supported by NIH grants R01AI063331 and R01DK091191 to G.N., research funds by the Israel Science Foundation (grant no. 768/11) to B.M., and funds to the Michigan Comprehensive Cancer Center Immunology Monitoring Core from the University of Michigan's Cancer Center Support Grant.

Author Contributions Y.H. and G.N. designed the research and wrote the manuscript. Y.H. conducted the experiments and analysed data with the help from M.Y.Z. and D.Y., and B.M. generated and provided critical material. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.N. (gabriel.nunez@umich.edu).

METHODS

Mice. *Nek7^{+/−}*, *Nlrp3^{−/−}*, *Asc^{−/−}* (also known as *Pycard^{−/−}*) *Casp1^{−/−}* *Casp11^{−/−}*, *Casp11^{−/−}* (also known as *Casp4^{−/−}*) mice on C57BL/6 background have been reported^{14,31–33}. NLRP3(R258W) mice were originally provided by Warren Strober (NIH). C57BL/6 mice were originally purchased from Jackson Laboratories and maintained in our facility. All animal studies were approved by the University of Michigan Committee on Use and Care of Animals.

Reagents. High-capacity streptavidin agarose resin was from Thermo Scientific (20359). S-protein agarose beads were from Novagen (69704). Biotin (B4501), nocodazole (M1404) and colchicine (C3915) were purchased from Sigma. LFM-A13 (1300) was purchased from Tocris. DOTAP liposomal transfection reagent was from Roche (11202375001). CytoTox 96 Non-Radioactive Cytotoxicity Assay Kit was purchased from Promega (G1780). LPS-B5 Ultrapure (tRL-pb5lps), LPS-SM Ultrapure (tRL-smlps), MSU (tRL-msu), CPPD (tRL-cppd), Nano-SiO₂ (tRL-sio) and poly(dA:dT)/lyovect (tRL-patc) were purchased from InvivoGen. Alum (77161) was from Thermo Fisher Scientific. ATP (A2383) was from Sigma. Nigericin (481990) was purchased from EMD Millipore. Silica (Min-U-Sil 5) was from US Silica. LLOMe was purchased from Chem-Impex. *Salmonella enterica* serovar Typhimurium strain SL1344 was a gift from D. Monack. Human NEK7 cDNA clone was purchased from the Harvard DNA Resource Core (HsCD00082815). Mouse *Nek7* cDNA clone was from Thermo Scientific (MM1013-202764053). Lentiviral expression vector pHIV-EGFP was from Addgene (21373). QuikChange II XL Site-Directed Mutagenesis Kit was from Agilent (20051). Antibodies for NEK7 (ab133514), NEK6 (ab133494) and NEK9 (ab138488) were purchased from Abcam. Mouse anti-NLRP3 (Cryo-2) was from AdipoGen2. Actin (A00730-200), HA tag (A01244-100) and Flag (DYKDDDDK) tag (A00187-200) antibodies were purchased from GenScript. Anti-α/β tubulin and anti-BTK antibodies were from Cell Signaling (2148, 3533). ASC antibody and caspase-1 antibody for the cleaved p20 of caspase-1 were generated in our laboratory.

Cell culture. BMDMs were generated by differentiating bone marrow progenitors from the tibia and femur for 7 days in 10% FCS IMDM (Gibco) supplemented with 30% L-cell supernatant, non-essential amino acids, sodium pyruvate and antibiotics (penicillin/streptomycin). iBMDMs were generated as previously described³⁴. iBMDMs were cultured in 10% FCS IMDM (Gibco) supplemented with non-essential amino acids, sodium pyruvate and antibiotics (penicillin/streptomycin). HEK293T cells (ATCC) were cultured on DMEM (Sigma) containing 10% FCS and antibiotics (penicillin/streptomycin). HEK293T cells were not authenticated or tested for mycoplasma contamination.

Purification of NLRP3-associated proteins. To search for interacting partners of NLRP3 in mouse macrophages, we generated a triple-tagged NLRP3 (NLRP3-SFP) that was fused with three tags in the C terminus: S-tag, Flag (for detection) and streptavidin-binding tag. This tagged NLRP3 can be used to purify NLRP3 protein without antibody-based immunoprecipitation. NLRP3-SFP was cloned into the pHIV-EGFP lentiviral expression vector (Addgene). Transduced immortalized *Nlrp3^{−/−}* macrophage clones were sorted by flow cytometry using green fluorescent protein (GFP) as a marker and screened for NLRP3 expression by immunoblotting. Clones with NLRP3 expression levels comparable to those found in wild-type iBMDMs stimulated with LPS were selected for further analysis. Macrophages were plated in 10-cm Petri dishes and primed with 200 ng ml^{−1} LPS for 4 h before stimulation with 5 mM ATP for 30 min. Cells were washed once with cold PBS after stimulation, and then lysed in ice-cold lysis buffer (50 mM Tris-HCl, pH 7.4, 2 mM EDTA, 150 mM NaCl, 0.5% Nonidet P-40, 1× EDTA-free Roche protease inhibitor cocktail) for 10 min at 4°C. Cell lysates were collected and spun down at 20,000g for 15 min at 4°C. The soluble fraction was incubated with streptavidin agarose beads for 2 h at 4°C. The beads were washed three times with lysis buffer. Bound proteins were eluted with 2 mM biotin in PBS, and then re-captured by incubation with S-protein agarose beads. The NLRP3-associated proteins were finally eluted in SDS sample buffer (10% glycerol, 60 mM Tris-HCl, pH 6.8, 2% SDS, 0.02% bromophenol blue and 5% β-mercaptoethanol), separated by 4–12% SDS-PAGE, followed by mass spectrometry. *Nlrp3^{−/−}* macrophages reconstituted with empty vector (pHIV-EGFP) were used as a control.

HEK293T cell transfection. HEK293T cells were plated into 6-well tissue culture plates (6.25×10^5 cells per well in 2 ml complete DMEM) overnight. Cells were transfected for 16 h with triple Flag-tagged full-length NLRP3 (0.5 µg), pyrin domain deletion mutant (Δpyrin; amino acids 94–1036, 0.5 µg), LRR deletion mutant (ΔLRR; amino acids 1–741, 1 µg), pyrin domain (amino acids 1–93, 0.5 µg), NOD domain (amino acids 220–536, 2.5 µg), LRR domain (amino acids 742–991, 1 µg), or co-transfected Flag-tagged full-length NLRP3 (0.5 µg) with NEK7-3×HA (0.5 µg) by Lipofectamine LTX (Invitrogen).

Immunoprecipitation. Cells were lysed in ice-cold lysis buffer (50 mM Tris, pH 7.4, 2 mM EDTA, 150 mM NaCl, 0.5% Nonidet P-40, 1× EDTA-free Roche protease inhibitor cocktail). Cell lysates were clarified by centrifugation (14,000g

at 4°C for 10 min. Pre-cleared cell lysates were incubated with anti-NEK7 (1:100), anti-NLRP3 (1:100), anti-Flag (1:200), anti-HA (1:200) antibody or control IgG at 4°C overnight. The proteins bound by antibody were pulled down by protein G beads and subjected to immunoblotting analysis.

Inflammasome activation and cytotoxicity assay. Macrophages were plated in 12-well plates at 1×10^6 cells per well. One day later, culture medium was replaced with 0.5 ml serum-free IMDM per well. Cells were then primed with 200 ng ml^{−1} Ultrapure LPS for 4 h, followed by stimulation with PBS (mock), 5 mM ATP (30 min), 5 µM nigericin (1 h), 0.5 µM gramicidin (1 h), poly(dA:dT) (2 µg ml^{−1}, 4 h), *Salmonella* (multiplicity of infection (m.o.i.) = 10, 1 h), LLOMe (2 µM, 6 h), silica (500 µg ml^{−1}, 6 h), Alum (250 µg ml^{−1}, 6 h), MSU (200 µg ml^{−1}), CPPD (200 µg ml^{−1}, 6 h) or Nano-SiO₂ (200 µg ml^{−1}, 6 h). After stimulation, supernatants and cell lysates were collected separately, or combined together for immunoblotting analysis. For non-canonical inflammasome activation, 1 µg of LPS-SM was packaged with 15 µl of DOTAP according to manufacturer's instructions. After 30 min incubation at room temperature, the mixtures were diluted into 1 ml Opti-MEM and added to each well of 6-well plate containing (LPS-B5)-primed macrophages (4 h priming). Four hours later, culture supernatants and cell lysates were collected for immunoblotting, and 48-well plates were used for IL-1β release and cytotoxicity assay.

Gene knockdown in primary macrophages. Lentiviral pLKO.1 plasmids targeting *Nek7*, *Nek6* and *Nek9* were purchased from Sigma. *Nek7* shRNA1 (TRC7RCN0000236005, targeting sequence: 5'-CCGGTGGAGTGCCGGTAG CGTTAAACTCGAGTTAACGCTACCGGCACCTCCATTGGT-3'), *Nek7* shRNA2 (TRCN0000236007, 5'-CCGGGAGCTACGACAGCTAGTTAA TCTCGAGATTAACAGCTGCTGAGCTCTTTTG-3'), *Nek6* shRNA (TRCN0000274723, 5'-CCGGAGGACAGTCAGTGAGGTTACTCGAGT AAACCTCACTGAACGTGCTTTTG-3') and *Nek9* (TRCN000027597, 5'-CCGGCGACAACATCATTGCCTACTACTCGAGTAGTAGGCAATGATGT TGTCGTTTT-3') were reported in this study. pLKO.1 scramble (Addgene 1864) was used as a negative control. Lentiviruses were packaged in HEK293FT cells and concentrated by Lenti-X concentrator (Clontech). Gene knockdown in primary macrophages was performed as described previously²⁰. Puromycin-selected macrophages were collected, counted and plated on day 8, and used for experiments on the next day. The knockdown efficiency was analysed by immunoblotting with antibodies against NEK7, NEK6 or NEK9.

Targeting of NEK7 by CRISPR-Cas9-mediated genome editing. Lentiviral CRISPR-Cas9 targeting guide RNA (gRNA) expressing vector (lentiCRISPRV2) was obtained from Addgene (52961). The NEK7 knockout target sequence used was 5'-TGAAAAACCGACCAAGCCC-3'. To generate a NEK7 knockout cell line, lentiCRISPRV2 stocks containing the target sequence 5'-TGAAAAACCGACCAAGCCC-3' were used to transduce iBMDMs. Three days later, single clones of puromycin-resistant cells were sorted by flow cytometry and clones with NEK7-deficiency were identified by immunoblotting with anti-NEK7 antibody.

Reconstitution of NEK7 in *Nek7^{−/−}* macrophages. Immortalized *Nek7^{−/−}* macrophages were transduced with lentivirus containing pHIV-NEK7-3×HA. After 3–4 days, transduced cells were sorted by flow cytometry using eGFP as a marker. Expression of reconstituted proteins was determined by immunoblotting.

Generation of chimaeras from fetal liver cells. Fetal livers were collected at day 14 of gestation from wild-type, *Nek7^{−/−}* or *Nlrp3^{−/−}* embryos. Wild-type and *Nek7^{−/−}* fetuses were genotyped with primers (wild-type, forward, 5'-AAGG CTTACTTGGTGACACACTGG-3'; reverse, 5'-CACCGTGCAGGTGACTC GAAC-3'; *Nek7^{−/−}*, forward, 5'-CCCTGGCGTTACCCAACCTTAATCGCCT TGC-3', reverse, 5'-TTCCCGTGGAAACAAACGGCGATTGAC-3'), which yielded a 400-base-pair (bp) wild-type DNA band and a 302-bp *Nek7* mutant DNA band after electrophoresis on 1.5% agarose gel in TBE buffer. Fetal livers were collected from fetuses and processed into single-cell suspensions using a 40 µm Nylon Cell Strainer (Falcon). Approximately 5×10^6 fetal liver cells in 100 µl PBS were administered by retro-orbital injection into 6-week-old gender-matched wild-type or *Nlrp3^{−/−}* recipient mice that had been lethally irradiated using X-ray (Phillips RT250, Kimtron Medical) with two doses of 5.40 Gy (total 10.8 Gy). Recipient mice were allocated randomly into experimental groups. Recipient animals received antibiotics (neomycin and polymyxin B) for 4 weeks after reconstitution. Eight weeks later, successful generation of chimaeras was confirmed by PCR-based analysis and animals were analysed (see above). The number of animals per group ($n=5$ –8) was chosen as the minimum likely required for conclusions of biological significance, established from previous experience.

ASC speck staining and ASC oligomer cross-linking. BMDMs were plated on an 8-well permanox chamber slide (Thermo Scientific, 177445) overnight. Cells were primed with 200 ng ml^{−1} LPS for 4 h, then stimulated with ATP or nigericin, transfected with poly(dA:dT), or infected with *Salmonella*. After stimulation, cells

were fixed with 4% paraformaldehyde, permeabilized with 0.1% Triton X-100, and the slides were blocked with PBS buffer containing 3% BSA. Cells were stained with anti-ASC antibody and Alexa-Fluor-488-conjugated secondary antibody. DAPI was used to stain nuclei. Cell images were taken using an Olympus Fluo-View 500 confocal microscope system.

For ASC oligomer cross-linking, cells were plated on 12-well plates and stimulated as indicated. Cells were lysed with PBS buffer containing 0.5% Triton X-100, and the cell lysates were centrifuged at 6,797g for 15 min at 4°C. Supernatants were transferred to new tubes (Triton-X-soluble fractions). The Triton-X-100-insoluble pellets were washed with PBS twice and then suspended in 200 µl PBS. The pellets were then cross-linked at room temperature for 30 min by adding 2 mM bis[sulfosuccinimidyl]suberate (BS³). The cross-linked pellets were spun down at 6,797g for 15 min and dissolved directly in SDS sample buffer.

Blue native PAGE and 2D PAGE. Blue native gel electrophoresis was performed using the Bis-Tris Native PAGE system as previously described^{20,35}. In brief, 2 × 10⁶ macrophages were re-plated in 6-well plates. On the second day, cells were primed with 200 ng ml⁻¹ LPS for 4 h, followed by stimulation with PBS (mock), 5 mM ATP (30 min) or 5 µM nigericin (1 h). Macrophages were washed once with cold PBS and then lysed in ice-cold native lysis buffer (20 mM Bis-tris, 500 mM ε-aminocaproic acid, 20 mM NaCl, 10% (w/v) glycerol, 0.5% digitonin, 0.5 mM Na₃VO₄, 1 mM PMSF, 0.5 mM NaF, 1 × EDTA-free Roche protease inhibitor cocktail, pH 7.0) for 15 min on ice. Cell lysates were clarified by centrifugation at 20,000g for 30 min at 4°C and analysed without further purification steps. Cell lysates were equalized after quantification of total protein using the BCA protein assay (Pierce), and then separated by 4–12% blue native PAGE. Native gels were incubated in 10% SDS solution for 5 min before transfer to PVDF membranes (Millipore), followed by conventional western blotting. In 2D PAGE, a gel slice of the natively resolved gel was placed in a dish containing 1 × Laemmli sample buffer (12.5 mM Tris-Cl (pH 6.8), 5% β-mercaptoethanol 4% (w/v) SDS, 0.1% bromophenol blue, 20% (v/v) glycerol) for 10 min, microwaved on high for 20 s, and rocked for another 15 min before loading the slice into a well 4–12% SDS-PAGE gel. Because the speck was mostly lost in the pellet fraction, the analyses of the NEK7–NLRP3 interaction probably excluded proteins in the speck.

Potassium efflux assay. Intracellular potassium measurements were performed as described previously⁸. In brief, macrophages were plated on 96-well plates 1 day

before experiments. Culture medium was thoroughly aspirated after stimulation. Cells were lysed with 3% ultrapure HNO₃. Intracellular potassium was then determined by inductively coupled plasma optical emission spectrometry with an Optima 2000 DV spectrometer (PerkinElmer Life Sciences) using yttrium as an internal standard.

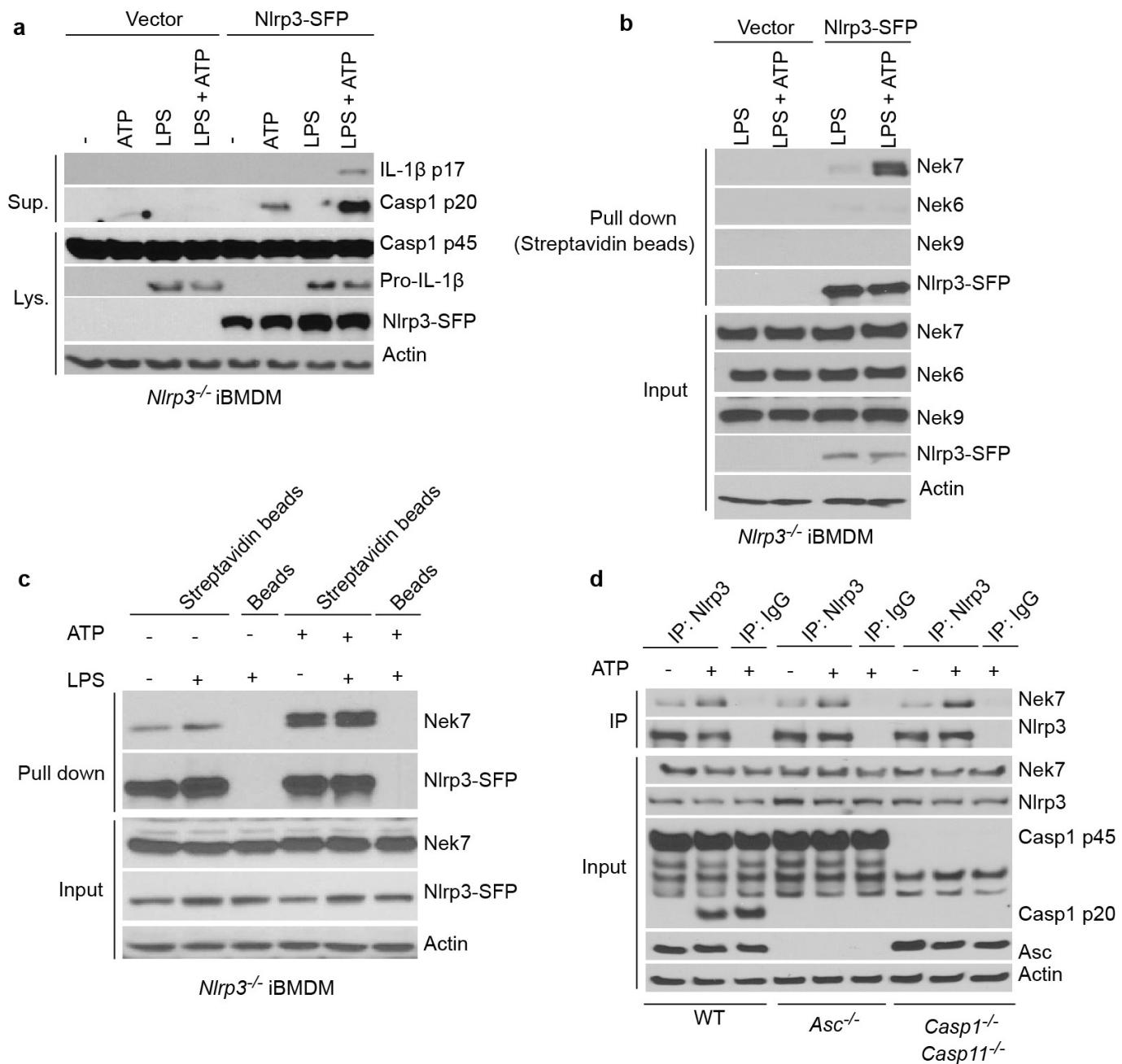
Cytokine measurements. Cytokines were measured with ELISA kits (R&D Systems).

Stimulation with endotoxin *in vivo*. Mice were injected intraperitoneally with 20 mg kg⁻¹ LPS (*Escherichia coli* 0111:B4; Sigma). The investigators were not blinded to allocation during experiments. The blood samples were collected 3 h later. Serum cytokines were measured by ELISA. The analysis was performed blindly by an independent researcher, and the experiments were not randomized.

MSU-induced mouse peritonitis. Chimaeric mice were injected intraperitoneally with 1 mg MSU dissolved in 0.5 ml sterile PBS. Mice were euthanized 6 h later and peritoneal cavities were flushed with 5 ml cold PBS. Peritoneal lavage fluids were collected and cytokines were measured by ELISA after concentration using an Amicon Ultra 10K filter from Millipore. The investigators were not blinded to allocation during experiments. Serum cytokines were measured by ELISA. The analysis was performed blindly by an independent researcher, and the experiments were not randomized.

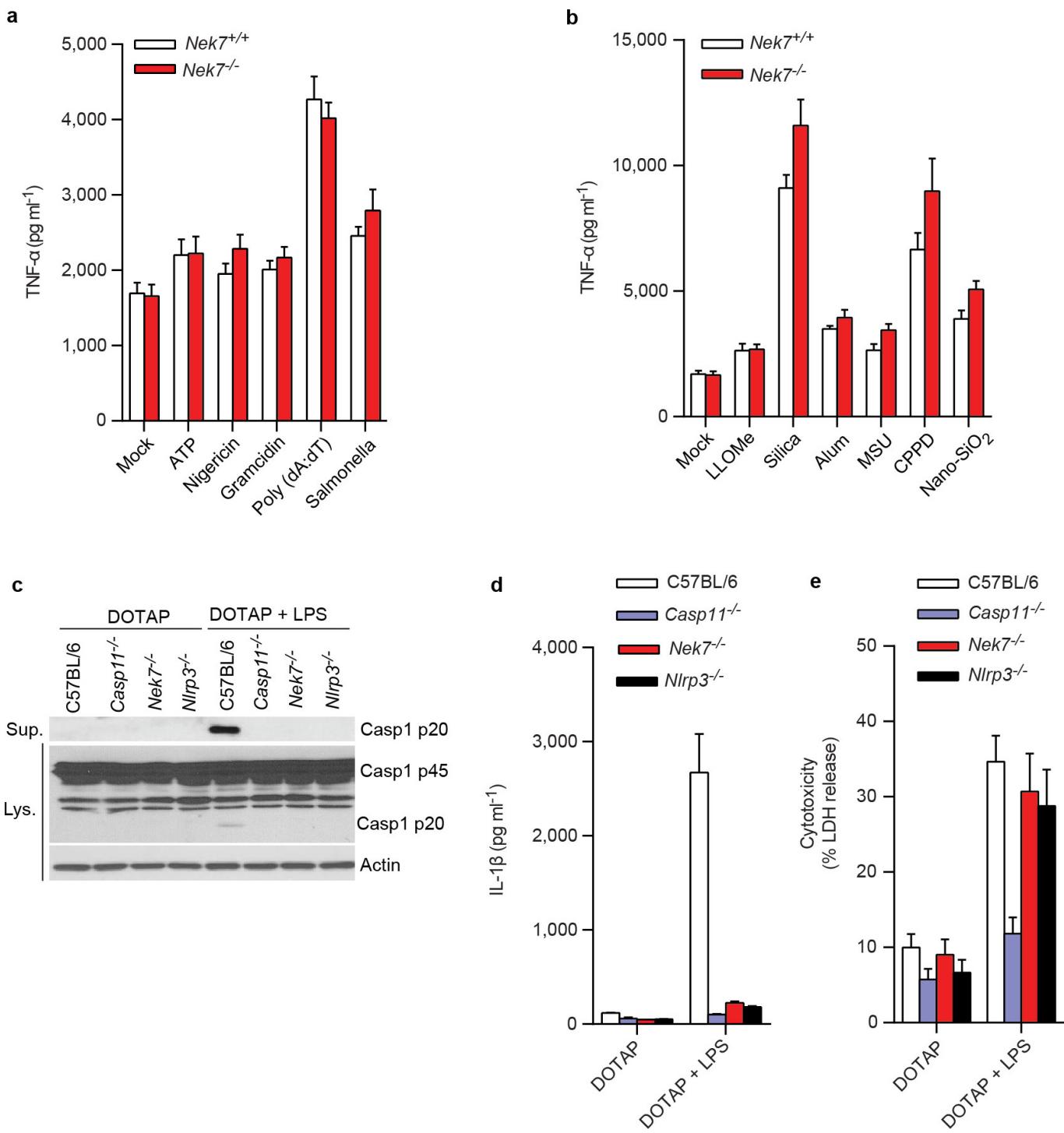
Statistical analysis. No statistical methods were used to predetermine sample size. All analyses were performed using GraphPad Prism. Differences were considered significant when *P* < 0.05.

31. Kanneganti, T. D. *et al.* Bacterial RNA and small antiviral compounds activate caspase-1 through cryopyrin/Nalp3. *Nature* **440**, 233–236 (2006).
32. Wang, S. *et al.* Murine caspase-11, an ICE-interacting protease, is essential for the activation of ICE. *Cell* **92**, 501–509 (1998).
33. He, Y., Franchi, L. & Nunez, G. TLR agonists stimulate Nlrp3-dependent IL-1beta production independently of the purinergic P2X7 receptor in dendritic cells and *in vivo*. *J. Immunol.* **190**, 334–339 (2013).
34. Blasi, E. *et al.* Selective immortalization of murine macrophages from fresh bone marrow by a raf/myc recombinant murine retrovirus. *Nature* **318**, 667–670 (1985).
35. Swamy, M., Siegers, G. M., Minguez, S., Wollscheid, B. & Schamel, W. W. Blue native polyacrylamide gel electrophoresis (BN-PAGE) for the identification and analysis of multiprotein complexes. *Sci. STKE* **2006**, pl4 (2006).



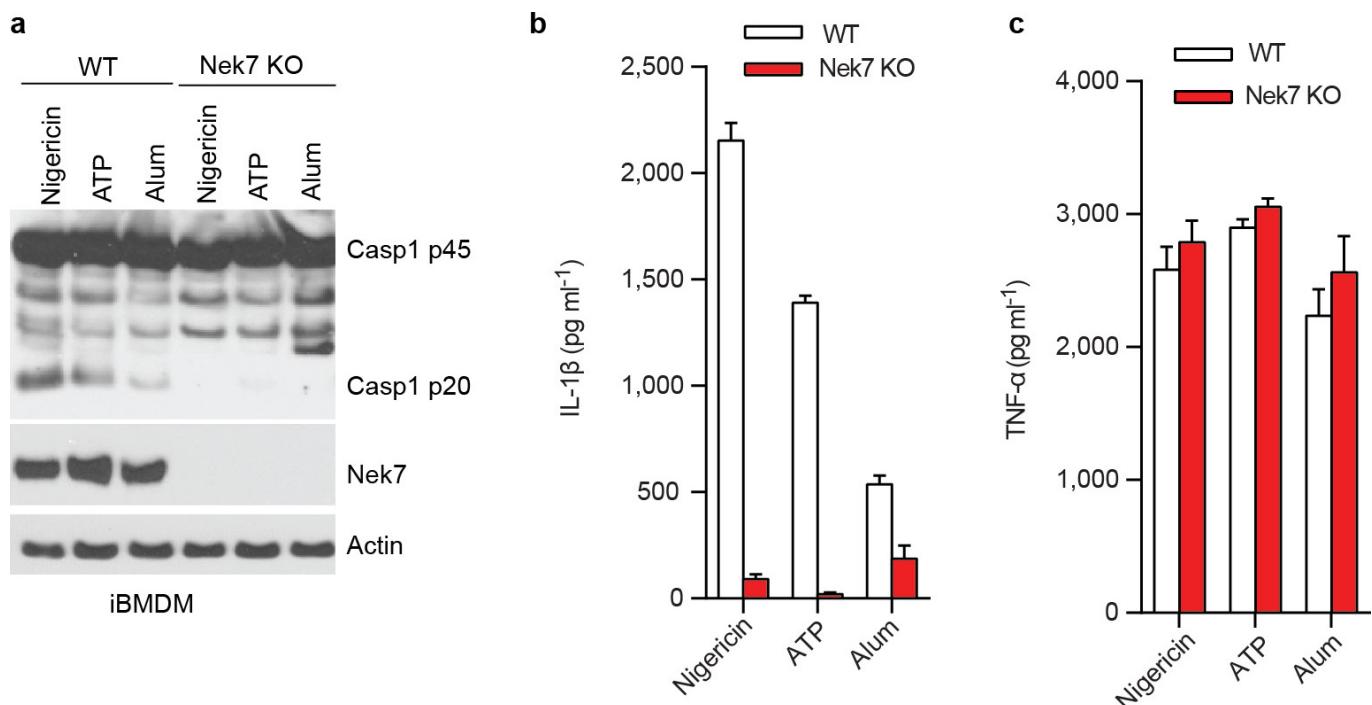
Extended Data Figure 1 | NEK7 interacts with NLRP3 upstream of ASC and caspase-1. **a**, Immortalized *Nlrp3*^{-/-} macrophages were infected with lentiviruses containing pHIV vector (Vector) or pHIV-NLRP3-SFP (NLRP3-SFP). Transduced macrophages were sorted by flow cytometry. Cells were treated as indicated. Supernatants and cell lysates were collected and analysed by immunoblotting with indicated antibodies. **b**, Immunoblots showing that NEK7, but not NEK6 or NEK9, interacts with NLRP3 in macrophages. NLRP3-associated proteins were pulled down from reconstituted *Nlrp3*^{-/-} macrophages by streptavidin

beads, and analysed by immunoblotting with indicated antibodies. **c**, ATP stimulation, but not LPS priming, greatly enhances the NEK7–NLRP3 interaction. **d**, LPS-primed BMDMs from wild-type, *Asc*^{-/-} and *Casp1*^{-/-} *Casp11*^{-/-} were left untreated or treated with 5 mM ATP for 30 min. NLRP3 protein complexes were immunoprecipitated with anti-NLRP3 antibody and analysed by immunoblotting. Data are representative of three (**a**, **b**) or two (**c**, **d**) independent experiments. See Supplementary Figs 2 and 3 for gel source data.



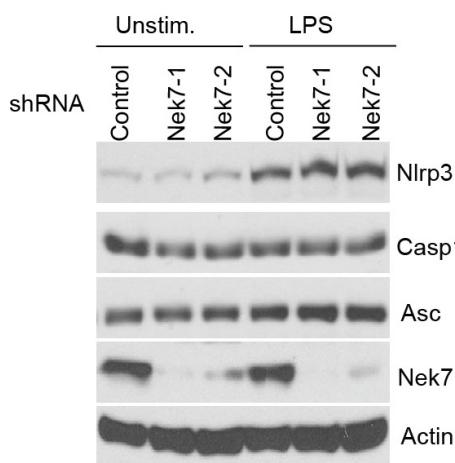
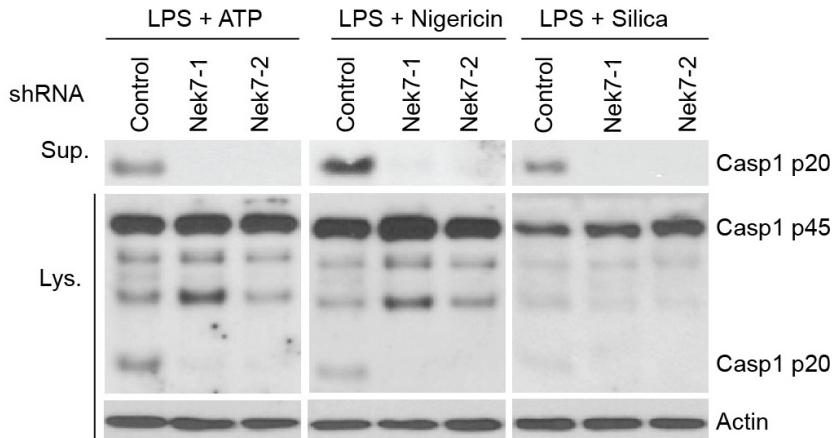
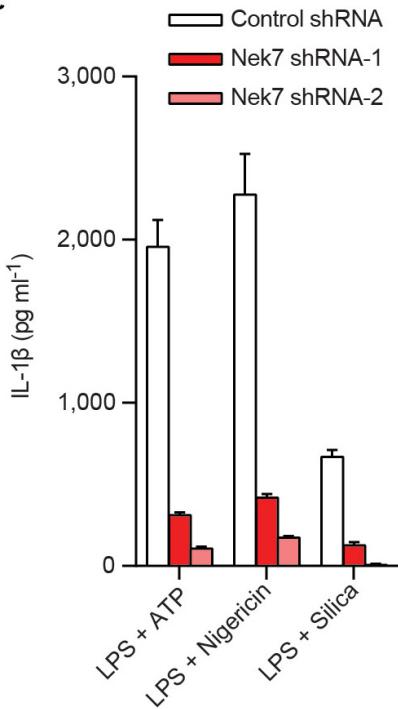
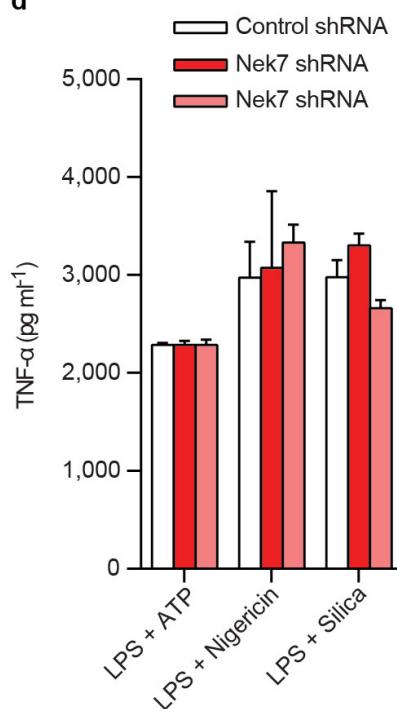
Extended Data Figure 2 | NEK7 deficiency does not affect TNF- α release, but abrogates NLRP3 activation induced by cytosolic LPS in macrophages. **a**, TNF- α secretion from LPS-primed BMDMs treated with ATP, nigericin, gramicidin, poly(dA:dT) or *Salmonella*. **b**, TNF- α secretion in LPS-primed BMDMs treated with LLOMe, silica, Alum, MSU, CPPD or nano-SiO₂. Mock represents macrophages primed with LPS without further stimulation. The amounts of TNF- α release in the absence of LPS

priming were below 50 pg ml⁻¹ in the supernatants. **c–e**, BMDMs from C57BL/6, *Casp11*^{-/-}, *Nek7*^{-/-} (fetal liver chimaeras) and *Nlrp3*^{-/-} mice were primed with LPS for 4 h before transfection with or without LPS with DOTAP. Four hours after transfection, caspase-1 activation (c), IL-1 β release (d) and cytotoxicity (e) were analysed. Graphs show the mean and s.d. of triplicate wells, and are representative of three independent experiments. See Supplementary Fig. 3 for gel source data.



Extended Data Figure 3 | NEK7 is required for activation of the NLRP3 inflammasome in macrophages. a–c, Defective NLRP3 inflammasome in NEK7-deficient macrophages. Wild-type and NEK7-deficient macrophages generated by CRISPR-Cas9 genome editing were primed with LPS, followed by stimulation with 5 μ M nigericin (1 h), 5 mM ATP (30 min) or 200 ng ml $^{-1}$ Alum (6 h). Cell lysates were collected and

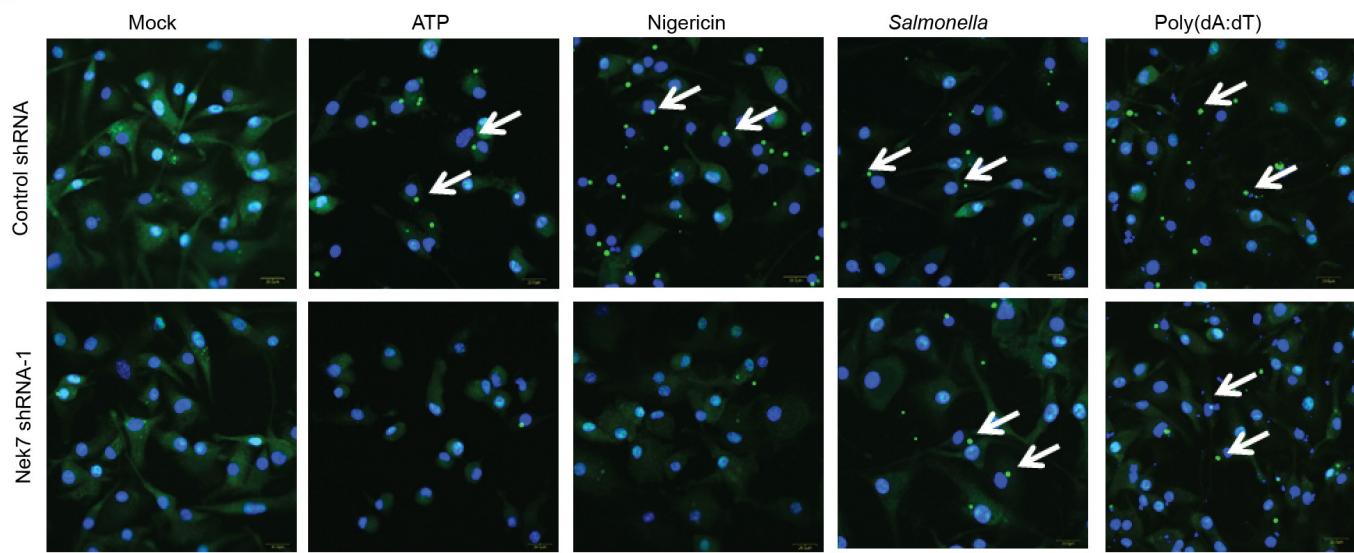
analysed for caspase-1 activation by immunoblotting (a). IL-1 β (b) and TNF- α (c) release were analysed by ELISA. ELISA data (b, c) show the mean and s.d. of triplicate wells. The amounts of TNF- α release in the absence of LPS priming were below 50 pg ml $^{-1}$ in the supernatants. Data are representative of three independent experiments. See Supplementary Fig. 3 for gel source data.

a**b****c****d**

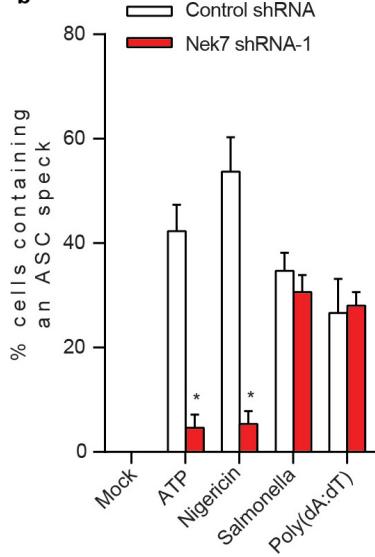
Extended Data Figure 4 | NEK7 is critical for activation of the NLRP3 inflammasome in primary macrophages. **a**, NEK7 is dispensable for NLRP3 induction and expression of the NLRP3 inflammasome components ASC and caspase-1. BMDMs were treated with control shRNA or *Nek7* shRNAs and selected by culture in $3\text{ }\mu\text{g ml}^{-1}$ puromycin. Puromycin-resistant macrophages were left unstimulated or stimulated with LPS for 4 h. Cell lysates were collected and analysed by immunoblotting. Actin is shown as a loading control. **b**, **c**, NEK7 depletion inhibits activation of the NLRP3 inflammasome. BMDMs treated with

control shRNA or *Nek7* shRNAs were primed with LPS, followed by stimulation with 5 mM ATP (30 min), 5 μM nigericin (1 h) or 500 ng ml^{-1} silica (4 h). Supernatants and cell lysates were collected after stimulation and analysed by immunoblotting for caspase-1 activation (**b**). IL-1 β (**c**) and TNF- α (**d**) release were analysed by ELISA. ELISA data (**c**, **d**) show the mean and s.d. of triplicate wells. The amounts of TNF- α release in the absence of LPS priming were below 50 pg ml^{-1} in the supernatants. Data are representative of three independent experiments. See Supplementary Fig. 3 for gel source data.

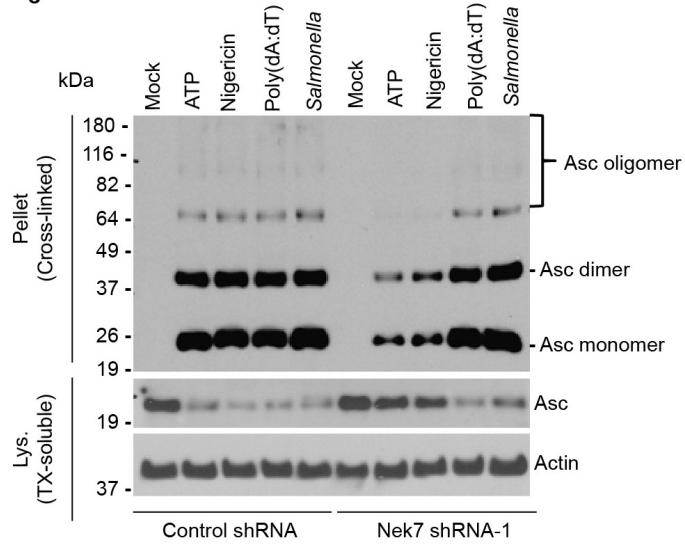
a



b



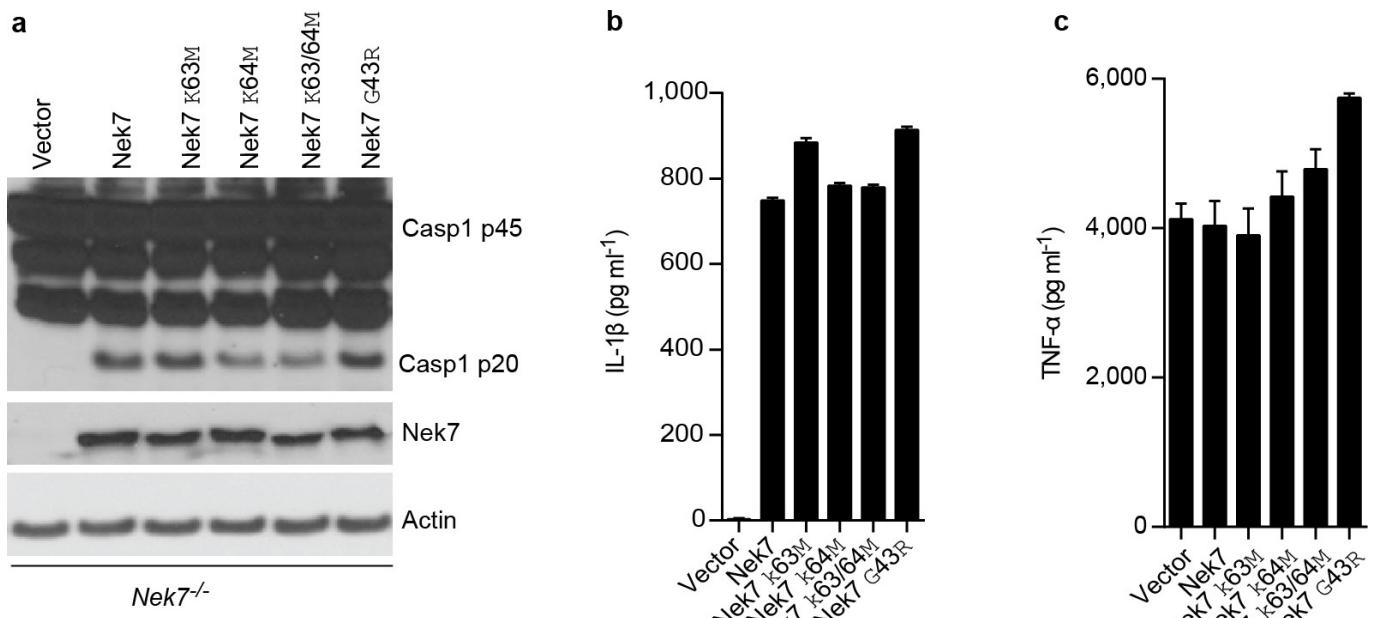
c



Extended Data Figure 5 | NEK7 is critical for NLRP3-mediated ASC oligomerization and speck formation in macrophages.

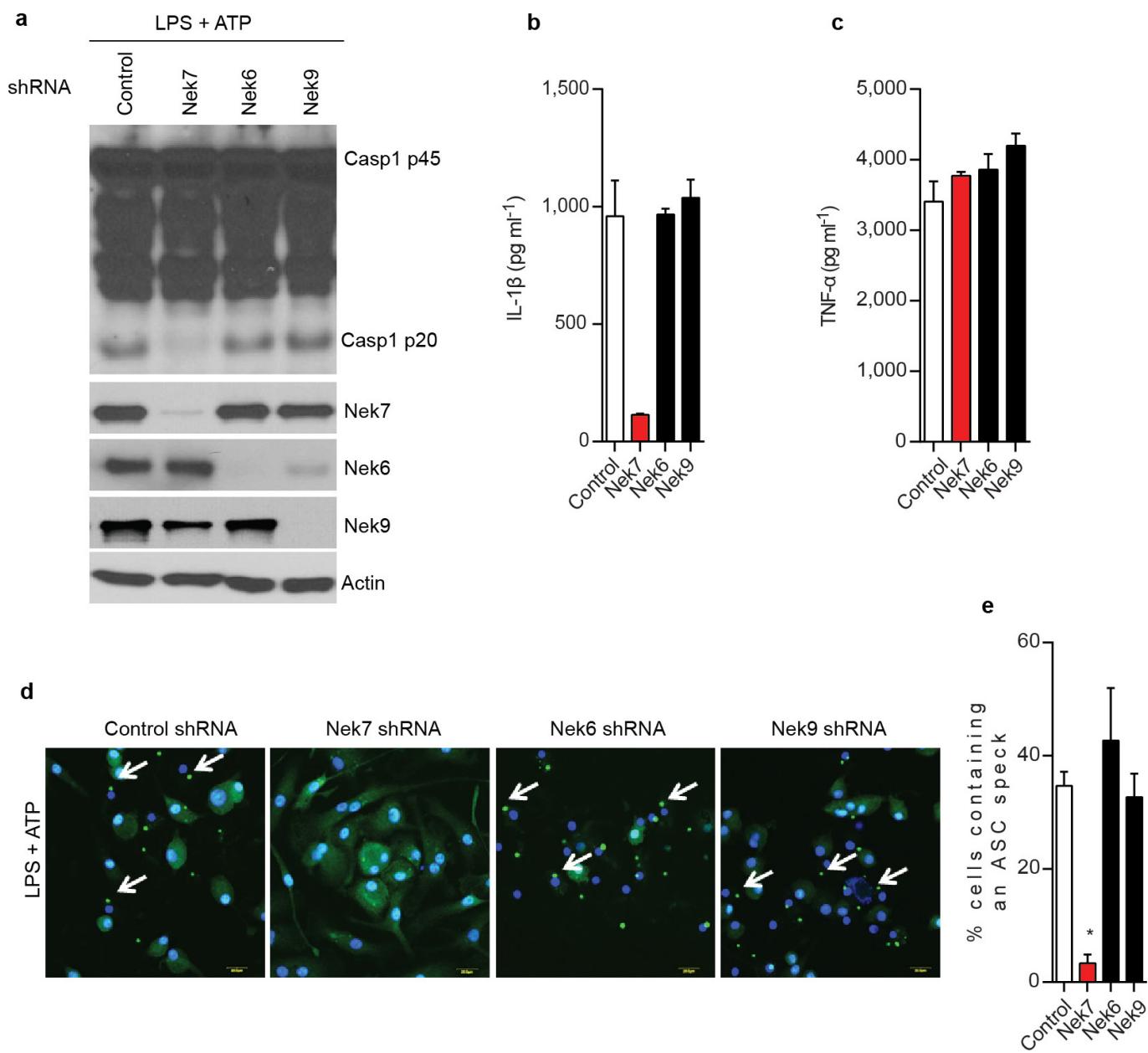
a, b, Representative confocal immunofluorescence images (a) and quantification of endogenous ASC specks (arrows) (b) in macrophages treated with control shRNA or *Nek7* shRNA. Macrophages were primed with LPS and then stimulated with ATP, nigericin, poly(dA:dT) or *Salmonella*. After stimulation, cells were fixed and stained for ASC (green). DAPI was used for staining nuclear DNA (blue). Data shown represent results from three combined independent experiments in

which more than 300 cells were counted in each experiment. Error bars indicate s.d. Original magnifications, $\times 400$. **c**, Immunoblots showing ASC oligomerization in control-shRNA-treated or *Nek7*-shRNA-treated macrophages. LPS-primed macrophages were treated as indicated. Cell lysates (Triton X-100 soluble) and BS³-crosslinked pellets (Triton X-100 insoluble) were analysed by immunoblotting. Mock represents macrophages primed with LPS without further stimulations. Results are representative of two independent experiments. $*P < 0.05$ (two-tailed Student's *t*-test). See Supplementary Fig. 3 for gel source data.



Extended Data Figure 6 | NEK7 kinase activity is dispensable for activation of the NLRP3 inflammasome in macrophages. a–c, Immortalized NEK7-deficient macrophages were infected with lentivirus containing empty vector or vector expressing wild-type or indicated mutant NEK7. Transduced macrophages were sorted by flow cytometry using GFP as a marker. Macrophages were plated and stimulated with

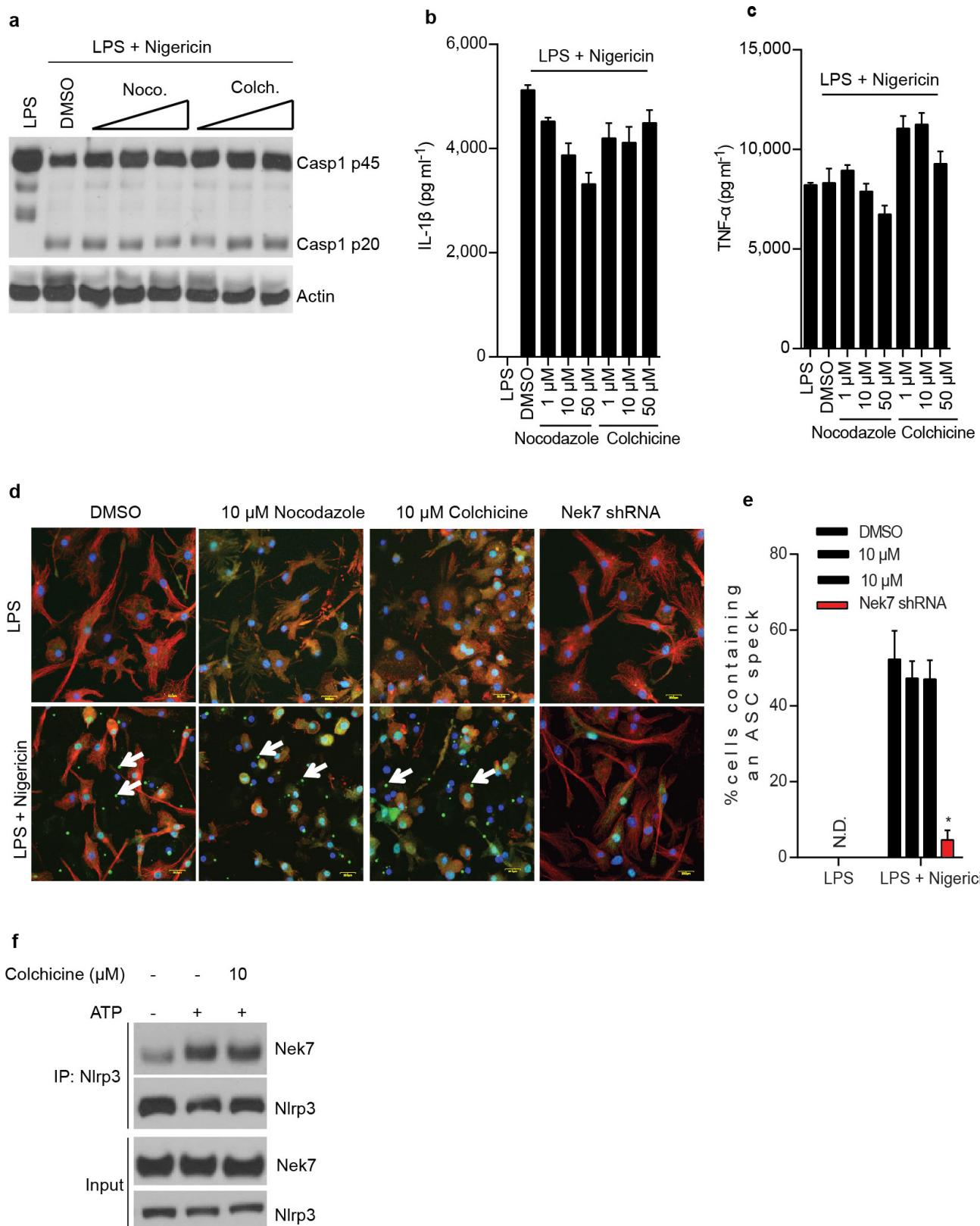
5 μ M nigericin for 1 h after LPS priming. Cell lysates were collected and analysed for caspase-1 activation by immunoblotting (a). IL-1 β (b) and TNF- α (c) release were analysed by ELISA. ELISA data (b, c) show the mean and s.d. of triplicate wells. Data are representative of three independent experiments. See Supplementary Fig. 3 for gel source data.



Extended Data Figure 7 | NEK6 and NEK9 are dispensable for activation of the NLRP3 inflammasome in macrophages.

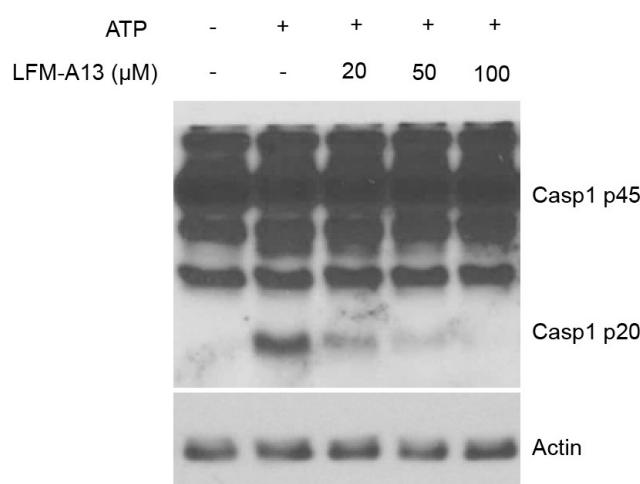
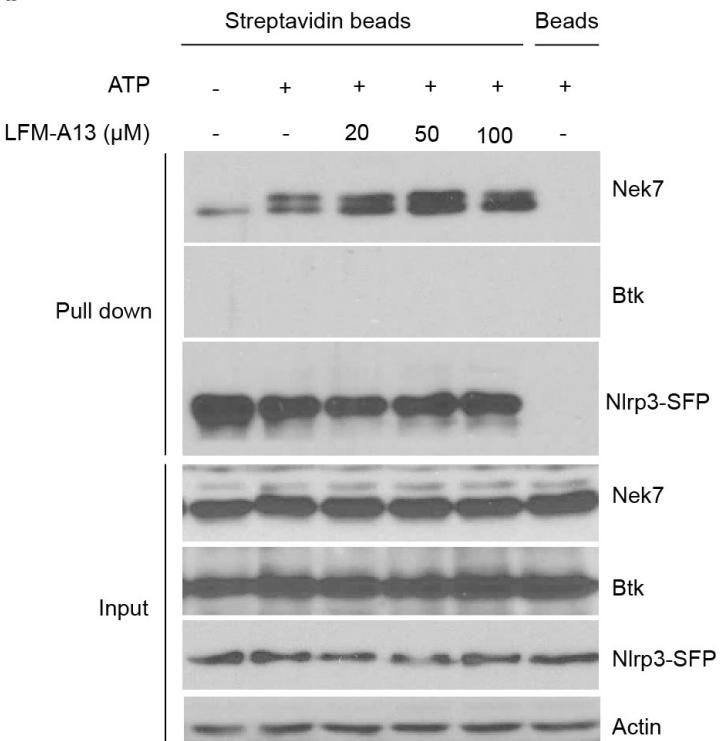
a–c, BMDMs were treated with control shRNA, Nek7 shRNA, Nek6 shRNA or Nek9 shRNA and selected by culture with $3\mu\text{g ml}^{-1}$ puromycin. Puromycin-resistant cells were plated and primed with LPS, followed by stimulation with 5 mM ATP for 30 min. Cell lysates were collected and analysed for caspase-1 activation by immunoblotting (a). IL-1 β (b) and TNF- α (c) release were analysed by ELISA. **d**, Representative confocal

immunofluorescence images (d) and quantification of endogenous ASC specks (arrows) (e) in macrophages. DAPI was used for staining nuclear DNA (blue). Quantification of ASC speck was from three combined independent experiments in which more than 300 cells were counted in each experiment. Error bars indicate s.d. Scale bars, 10 μm . ELISA data (b, c) show the mean and s.d. of triplicate wells. Data are representative of three independent experiments. * $P < 0.05$ (two-tailed Student's *t*-test). See Supplementary Fig. 3 for gel source data.



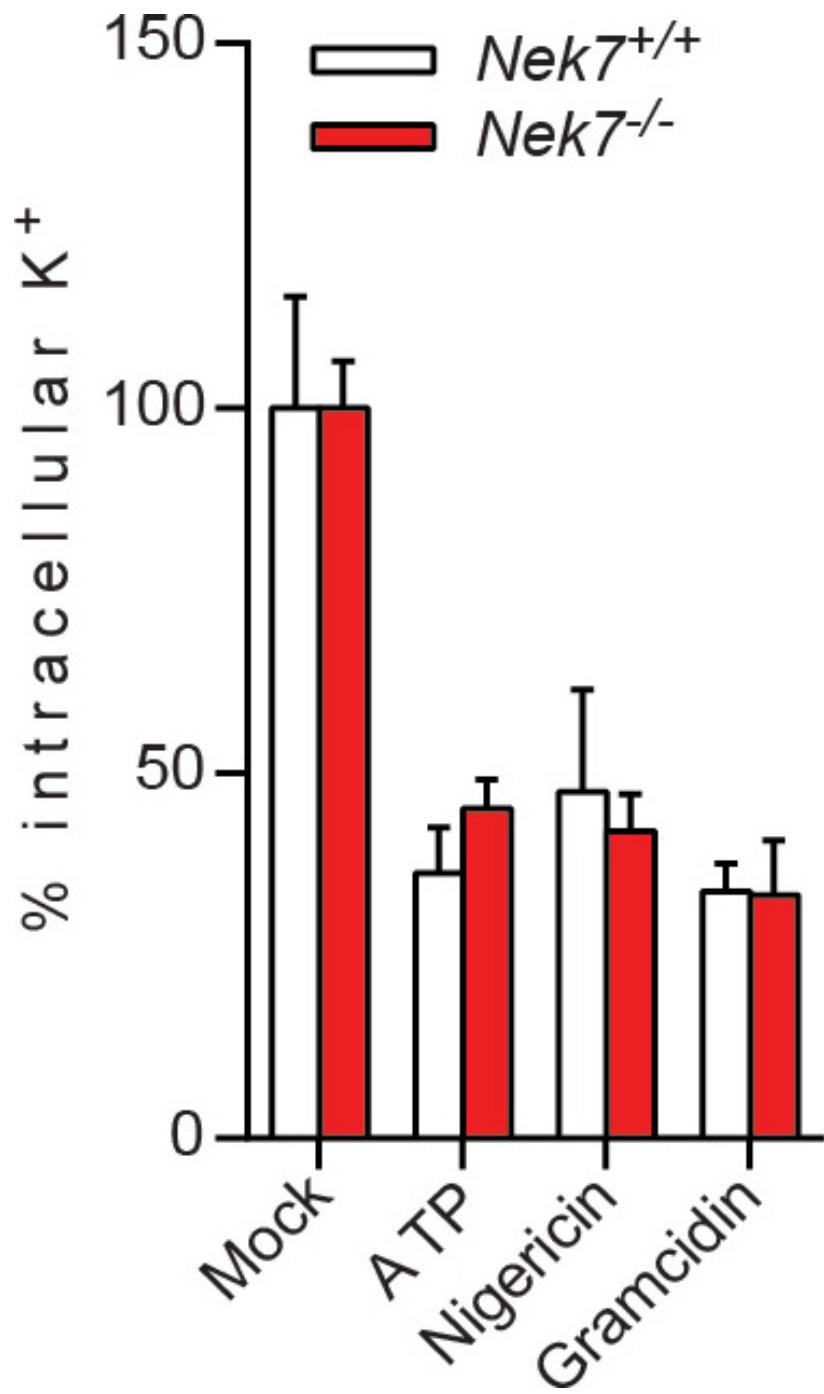
Extended Data Figure 8 | Microtubule depolymerization does not inhibit the activation of the NLRP3 inflammasome and the NEK7–NLRP3 interaction in macrophages. **a–e**, LPS-primed BMDMs were pretreated with vehicle control (DMSO), nocodazole (1, 10 or 50 μ M) or colchicine (1, 10 or 50 μ M) for 1 h before stimulation with 5 μ M nigericin. Cell lysates were collected and analysed for caspase-1 activation by immunoblotting (**a**). IL-1 β (**b**) and TNF- α (**c**) release were analysed by ELISA. Representative confocal immunofluorescence images (**d**) and quantification (**e**) of endogenous ASC specks (arrows) in macrophages treated as indicated.

Microtubules were stained with anti- α/β tubulin antibody (red). DAPI was used for staining nuclear DNA (blue). Data shown in **e** represent results from three combined independent experiments in which more than 300 cells were counted in each experiment. Error bars indicate s.d. Scale bars, 10 μ m. **f**, The NEK7–NLRP3 interaction in untreated and treated macrophages was analysed by immunoprecipitation and immunoblotting. ELISA data (**b**, **c**) show the mean and s.d. of triplicate wells. Data are representative of three independent experiments. * $P < 0.05$ (two-tailed Student's *t*-test). See Supplementary Fig. 3 for gel source data.

a**b**

Extended Data Figure 9 | Inhibition of BTK does not inhibit the NEK7–NLRP3 interaction in macrophages. **a, b,** Reconstituted immortalized *Nlrp3*^{−/−} macrophages were primed with LPS for 4 h. Cells were left untreated or treated with indicated concentrations of BTK inhibitor LFM-A13 for 30 min before ATP stimulation. Cell lysates were collected and analysed for caspase-1 activation by immunoblotting (a). Cell lysates

were collected and subjected to pull-down with streptavidin beads, and the precipitated protein complex was analysed by immunoblotting with indicated antibodies (b). Agarose beads were used as a control. Actin is shown as a loading control. Data are representative of two independent experiments. See Supplementary Fig. 3 for gel source data.



Extended Data Figure 10 | NEK7-deficiency does not inhibit potassium efflux induced by NLRP3 stimuli in macrophages. LPS-primed *Nek7*^{+/+} and *Nek7*^{-/-} BMDMs were stimulated as indicated. The intracellular potassium in each condition was analysed. Mock represents macrophages

primed with LPS without further stimulation. Graphs show the mean and s.d. of four technical replicates and are representative of two independent experiments.

Cryo-EM reveals a novel octameric integrase structure for betaretroviral intasome function

Allison Balandras-Colas¹, Monica Brown², Nicola J. Cook³, Tamaria G. Dewdney¹, Borries Demeler⁴, Peter Cherepanov^{3,5}, Dmitry Lyumkis² & Alan N. Engelman¹

Retroviral integrase catalyses the integration of viral DNA into host target DNA, which is an essential step in the life cycle of all retroviruses¹. Previous structural characterization of integrase-viral DNA complexes, or intasomes, from the spumavirus prototype foamy virus revealed a functional integrase tetramer^{2–5}, and it is generally believed that intasomes derived from other retroviral genera use tetrameric integrase^{6–9}. However, the intasomes of orthoretroviruses, which include all known pathogenic species, have not been characterized structurally. Here, using single-particle cryo-electron microscopy and X-ray crystallography, we determine an unexpected octameric integrase architecture for the intasome of the betaretrovirus mouse mammary tumour virus. The structure is composed of two core integrase dimers, which interact with the viral DNA ends and structurally mimic the integrase tetramer of prototype foamy virus, and two flanking integrase dimers that engage the core structure via their integrase carboxy-terminal domains. Contrary to the belief that tetrameric integrase components are sufficient to catalyse integration, the flanking integrase dimers were necessary for mouse mammary tumour virus integrase activity. The integrase octamer solves a conundrum for betaretroviruses as well as alpharetroviruses by providing critical carboxy-terminal domains to the intasome core that cannot be provided in *cis* because of evolutionarily restrictive catalytic core domain–carboxy-terminal domain linker regions. The octameric architecture of the intasome of mouse mammary tumour virus provides new insight into the structural basis of retroviral DNA integration.

Mouse mammary tumour virus (MMTV) intasomes were assembled from integrase (IN) and viral DNA (vDNA) components by differential salt dialysis, akin to the strategy used for prototype foamy virus (PFV) intasomes². Fractionation of assembly reactions by size-exclusion chromatography (SEC) revealed a higher-order species with a distinct elution profile from those of IN and vDNA (Fig. 1a). To address biological relevance, strand transfer reactions were conducted with supercoiled plasmid target DNA (tDNA) to monitor the concerted integration of two vDNA ends¹⁰ (Fig. 1b). The SEC-purified complexes catalysed efficient concerted integration activity, which was inhibited by the IN strand transfer inhibitor raltegravir (Fig. 1c). The sequencing of concerted integration products excised from agarose gels revealed that most contained 6 base pair (bp) target site duplications flanking the integrated vDNA ends, which are known to occur during MMTV infection¹¹ (Fig. 1d). To address the specificity of complex formation, the invariant CA dinucleotide, which is essential for IN catalysis^{12,13}, was mutated to GT in the vDNA substrate. As the mutant vDNA failed to support complex formation (data not shown), we conclude that the higher-order species identified by SEC are bona fide MMTV intasomes. We note that divalent metal ion was a critical cofactor for MMTV intasome formation. On the basis of prior reports that Ca^{2+} promoted

the assembly of active HIV-1 IN–vDNA complexes but was unable to support IN catalysis¹⁴, it was used here for intasome assembly.

To determine the MMTV intasome structure, single-particle cryo-electron microscopy (cryo-EM) data was collected on a Titan Krios microscope equipped with a Gatan K2 direct detector. Computational processing of the data revealed the most stable structural conformation of the complex, which was refined to $\sim 5\text{--}6\text{ \AA}$ for different regions of the map (Fig. 2a and Extended Data Figs 1 and 2). The MMTV intasome is composed of central core density as well as flanking density regions that are conformationally mobile compared with the intasome core (Extended Data Fig. 3). Restricting data refinement to the core density region accordingly increased the resolution for the

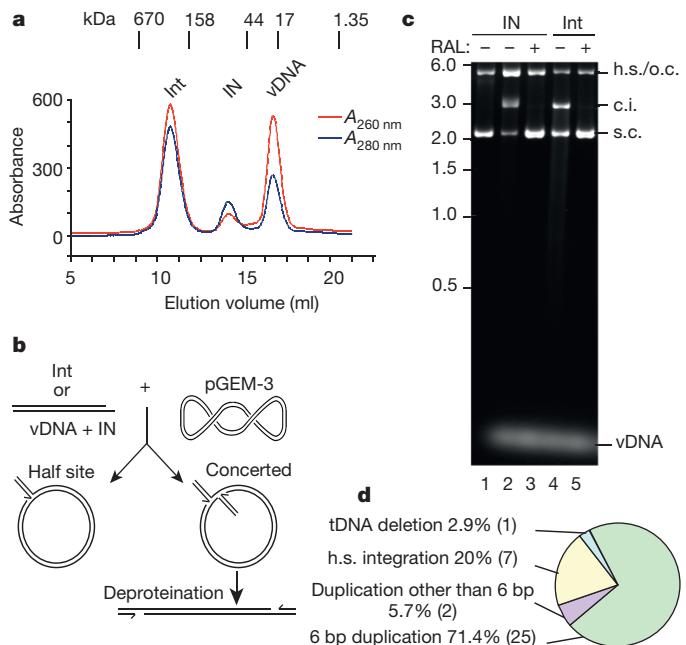


Figure 1 | MMTV intasome (Int) characterization. **a**, Purification by SEC. Elution positions of mass standards in kilodaltons are indicated. **b**, Integration assay schematic. Intasome or IN plus vDNA was reacted with supercoiled tDNA, which can yield half-site (h.s.) or concerted integration (c.i.) products. **c**, Ethidium bromide stained agarose gel. Reactions shown in lanes 1–3 were initiated with IN; vDNA was omitted from lane 1. Raltegravir (RAL) was included as indicated. Lanes 4 and 5, intasome reactions. Migration positions of half-site products that co-migrate with open circular (o.c.) tDNA, concerted integration products, supercoiled (s.c.) tDNA and mass standards in kilobases are indicated. For gel source data, see Supplementary Fig. 1. **d**, Sequenced intasome-mediated concerted integration products ($n=35$).

¹Department of Cancer Immunology and Virology, Dana-Farber Cancer Institute and Department of Medicine, Harvard Medical School, 450 Brookline Avenue, Boston, Massachusetts 02215, USA.

²Laboratory of Genetics and Helmsley Center for Genomic Medicine, The Salk Institute for Biological Studies, 10010 N Torrey Pines Road, La Jolla, California 92037, USA. ³Clare Hall Laboratories, The Francis Crick Institute, Blanche Lane, South Mimms, Potters Bar, Hertfordshire EN6 3LD, UK. ⁴Department of Biochemistry, The University of Texas Health Science Center at San Antonio, 7703 Floyd Curl Drive, San Antonio, Texas 78229, USA. ⁵Division of Medicine, Imperial College London, St. Mary's Campus, Norfolk Place, London W2 1PG, UK.

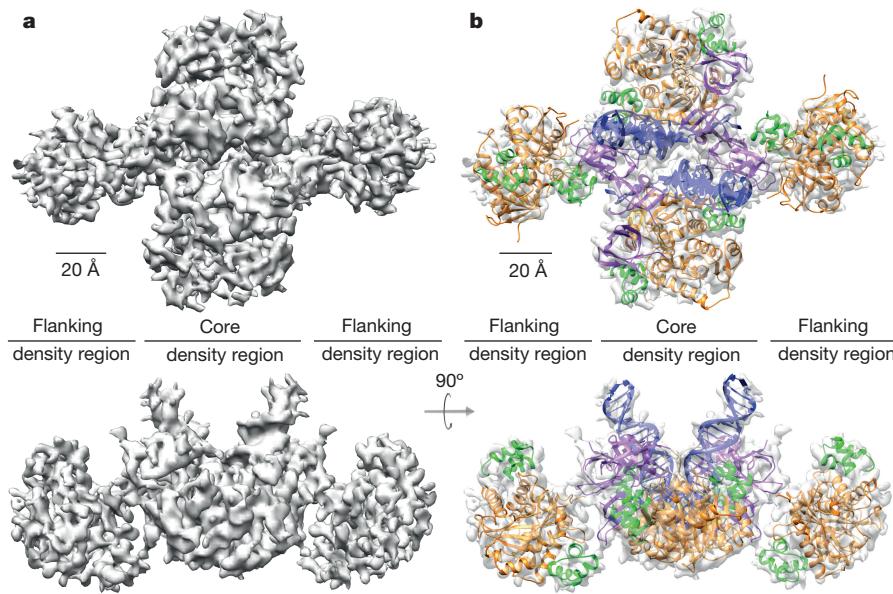


Figure 2 | Cryo-EM structure of the MMTV intasome. a, Top view (upper) of the cryo-EM map; the lower view is rotated by 90°. Core density and flanking density regions are indicated. **b,** Individual domain crystal

central portion of the structure to ~ 4 Å for the best-resolved regions (Extended Data Fig. 2d).

Each IN monomer is composed of an amino (N)-terminal domain (NTD), a catalytic core domain (CCD) and a carboxy (C)-terminal domain (CTD) (Extended Data Fig. 4a), and the map was sufficiently detailed to readily assign these domains to their corresponding cryo-EM densities. Given a lack of MMTV IN structures, the different protein domains were crystallized as IN_{CCD} , IN_{CTD} and $\text{IN}_{\text{NTD-CCD}}$ fragments, and these structures were refined to 1.7 Å, 1.5 Å and 2.7 Å resolution, respectively (Extended Data Fig. 5 and Extended Data Table 1). MMTV DNA was modelled using PFV intasome DNA coordinates and by extending the modelled fragment by 3 bp in the region distal from the IN active sites to account for the different vDNA lengths. Using rigid-body docking, the two vDNAs and eight NTDs, CCDs and CTDs were unambiguously positioned into the cryo-EM map (Fig. 2b). Rosetta^{15–17} was used to refine the X-ray structures and vDNA, and to build a subset of interdomain linker regions to best fit the density within the intasome core region (Extended Data Fig. 6 and Supplementary Videos 1–5). The resulting model revealed two molecules of vDNA and eight MMTV INs arranged as four IN dimers (Fig. 3a). Two catalytic IN dimers A and B are positioned in the core region in close contact to the vDNAs, whereas supportive IN dimers C and D locate to the flanking density regions, donating their CTDs to the core. Flexible linkers connect the IN domains, and the NTD–CCD linker, which is contracted in most IN protomers, extends in IN_1 and IN_3 to donate these NTDs in *trans* to opposing CCDs (Fig. 3a and Extended Data Fig. 6e). Sedimentation velocity centrifugation indicated the molecular mass of active MMTV intasomes as 302.1 kDa, which is fully consistent with the octameric IN structure (calculated $\text{IN}_8\text{--vDNA}_2 = 313.6$ kDa; Extended Data Fig. 4b).

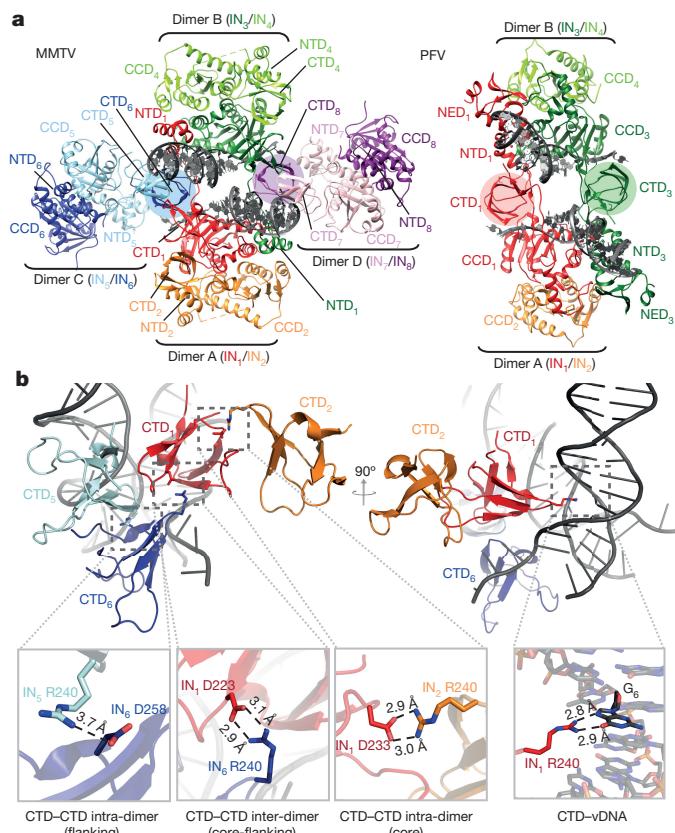
The structures of the MMTV and PFV intasomes were compared to ascertain aspects of the new structure important for DNA recombination (Fig. 3a). The PFV intasome is composed of two IN dimers A and B, with the inner protomers of each dimer (IN_1 and IN_3 ; red and green in Fig. 3a) adopting extended conformations². The NTDs and CTDs of the outer IN protomers (chartreuse (light green) and orange in Fig. 3a) are unseen in PFV intasome density maps. The architecture in the core density region of the MMTV intasome is strikingly similar to the PFV structure.

structures (NTD, green; CCD, orange; CTD, purple) and vDNA (blue) model fitted by rigid body docking. Rulers demarcate 20 Å.

For example, the positions of the CCDs and NTDs that contact vDNA (red IN_1 and green IN_3 in Fig. 3a) are analogous. The two remaining NTDs in the core region stabilize the CCD dimer interface in an arrangement identical to that seen in the $\text{IN}_{\text{NTD-CCD}}$ crystal structure (Extended Data Figs 5d and 6e). Both flanking density regions contain a CCD dimer that is also stabilized on each side by NTDs, mimicking the CCD dimer arrangements found in the core density region.

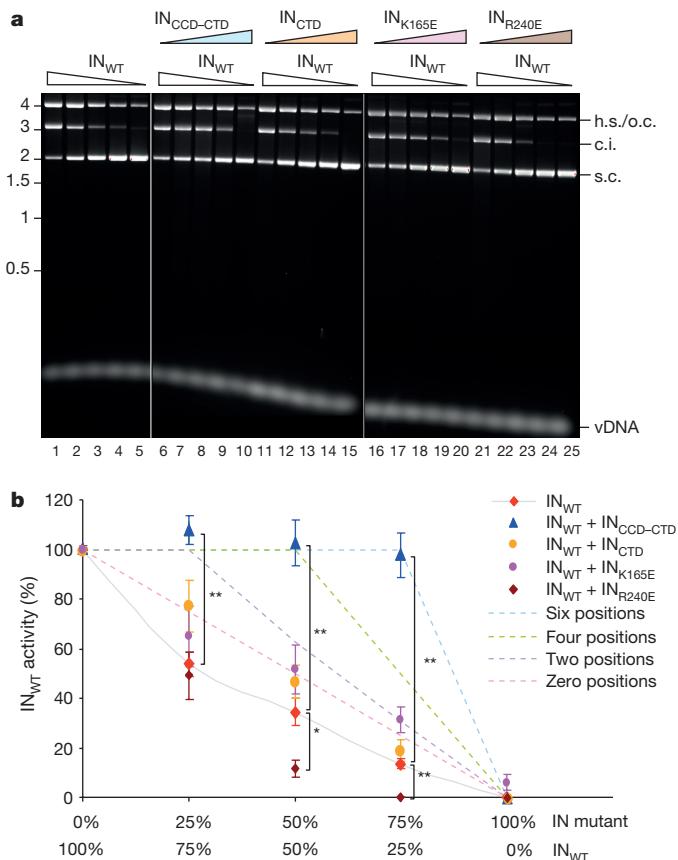
The arrangements of the CTDs differ dramatically between the MMTV and PFV structures, with MMTV IN residue Arg240 mediating several key contacts. For example, core protomer IN_1 Arg240 interacts with vDNA while IN_2 Arg240 interacts with IN_1 Asp233 (Fig. 3b). Flanking protomer IN_5 Arg240 engages its IN_6 neighbour whereas IN_6 Arg240 mediates an inter-dimeric interaction with core protomer IN_1 Asp223, docking the flanking IN dimer to the core structure (Fig. 3b). To test the functionality of the flanking IN dimers, complementation assays were performed by varying ratios of wild-type (IN_{WT}) and mutant IN proteins in strand transfer reactions. Similar strategies were used previously to dissect the division of labour within multimeric complexes of retroviral IN^{18–21} as well as the related bacteriophage Mu transpososome²².

IN_{R240E} , which like IN_{WT} purified as a dimer (Extended Data Fig. 7), was defective for strand transfer activity (Fig. 4a). To assess the functionality of Arg240-mediated IN–IN interactions, we compared IN_{R240E} with IN_{K165E} , which carries a change that uniquely disrupts IN–vDNA binding^{2,23}. In concordance with its inability to assume the roles of inner IN_1 and IN_3 subunits of the core tetramer, IN_{K165E} mildly stimulated the activity of limited IN_{WT} protein (Fig. 4b), presumably providing a source for other IN subunits within the functional complex. IN_{R240E} by contrast potently inhibited IN_{WT} function, confirming the importance of Arg240-mediated protein–protein interactions for MMTV IN activity. Two deletion mutant constructs, $\text{IN}_{\text{CCD-CTD}}$ and IN_{CTD} , which purified as dimers and monomers, respectively (Extended Data Fig. 7), were additionally analysed. The reaction composed of 75% $\text{IN}_{\text{CCD-CTD}}$ supported near IN_{WT} activity, indicating that this mutant could function when present in up to six of eight octamer positions. This finding strongly supports flanking IN dimer functionality, as the absence of the NTD would likewise preclude $\text{IN}_{\text{CCD-CTD}}$ from assuming intasome core positions 1 and 3. As the IN_{CTD} response



curve overlaid that predicted for non-functional protein, we moreover conclude that CCD-mediated dimerization is critical for flanking IN CTD function (Fig. 4).

Analysis of IN primary sequences suggests an explanation for the octameric arrangement of IN within the MMTV intasome when an IN tetramer suffices for PFV integration. Whereas fifty-residue CCD–CTD linkers afford the positioning of inner PFV IN CTDs for vDNA and tDNA engagement^{2,3}, the analogous eight-amino-acid MMTV linker is simply too short to accomplish the task (Extended Data Fig. 8a). MMTV has accordingly evolved to employ flanking IN dimers to nestle CTDs into the core intasome structure to provide essential CTD function in *trans* for integration. As flanking IN dimer CTDs 6 and 8 structurally mimic the PFV domains (Fig. 3a and Extended Data Fig. 8a), we presume these CTDs will engage tDNA during MMTV integration. Extending our analysis to other retroviruses indicates that in addition to the spumaviruses, IN tetramers could suffice for gamma- and epsilonretroviral intasome activity, while an IN octamer will be required to catalyse alpharetrovirus in addition to betaretrovirus integration (Extended Data Fig. 8b). We note that an octameric IN architecture for the alpharetrovirus Rous sarcoma virus intasome has recently been independently determined²⁴. Whereas most studies have highlighted a tetramer as the IN species that catalyses concerted HIV-1 integration^{9,25,26}, others have implicated a role for



octameric IN^{27,28}. Given the intermediary length of lentiviral IN CCD–CTD linker regions (Extended Data Fig. 8b), the higher-order nature of IN in active HIV-1 intasomes may need to be re-evaluated.

PFV IN, which cleaves tDNA phosphodiester bonds separated by 4 bp, preferentially integrates into flexible sequences, whereas MMTV and Rous sarcoma virus IN, which cleave tDNA with 6 bp stutters, are relatively unconstrained by tDNA flexibility^{3,29}. Superposition of the MMTV and PFV intasome structures revealed that the two sets of catalytic IN active sites almost perfectly aligned (Extended Data Fig. 8c). The same practical spacing of IN active sites therefore catalyses PFV and MMTV integration into sharply bent versus relatively non-deformed tDNA, respectively (Extended Data Fig. 8d). Owing to their positions in the structure, we note that the flanking IN dimers dramatically expand the potential contact area with tDNA, which is likely to have consequences for the docking of alpha- and betaretroviral intasomes to host chromatin.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 August; accepted 30 December 2015.

1. Craigie, R. & Bushman, F. D. HIV DNA integration. *Cold Spring Harb. Perspect. Med.* **2**, a006890 (2012).
2. Hare, S., Gupta, S. S., Valkov, E., Engelman, A. & Cherepanov, P. Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature* **464**, 232–236 (2010).
3. Maertens, G. N., Hare, S. & Cherepanov, P. The mechanism of retroviral integration from X-ray structures of its key intermediates. *Nature* **468**, 326–329 (2010).
4. Hare, S., Maertens, G. N. & Cherepanov, P. 3'-Processing and strand transfer catalysed by retroviral integrase *in crystallo*. *EMBO J.* **31**, 3020–3028 (2012).
5. Maskell, D. P. *et al.* Structural basis for retroviral integration into nucleosomes. *Nature* **523**, 366–369 (2015).
6. Yang, Z. N., Mueser, T. C., Bushman, F. D. & Hyde, C. C. Crystal structure of an active two-domain derivative of Rous sarcoma virus integrase. *J. Mol. Biol.* **296**, 535–548 (2000).
7. Wang, J. Y., Ling, H., Yang, W. & Craigie, R. Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein. *EMBO J.* **20**, 7333–7343 (2001).
8. Bao, K. K. *et al.* Functional oligomeric state of avian sarcoma virus integrase. *J. Biol. Chem.* **278**, 1323–1327 (2003).
9. Li, M., Mizuuchi, M., Burke, T. R., Jr & Craigie, R. Retroviral DNA integration: reaction pathway and critical intermediates. *EMBO J.* **25**, 1295–1304 (2006).
10. Hare, S. *et al.* A novel co-crystal structure affords the design of gain-of-function lentiviral integrase mutants in the presence of modified PSIP1/LEDGF/p75. *PLoS Pathog.* **5**, e1000259 (2009).
11. Majors, J. E. & Varmus, H. E. Nucleotide sequences at host-proviral junctions for mouse mammary tumour virus. *Nature* **289**, 253–258 (1981).
12. Roth, M. J., Schwartzberg, P. L. & Goff, S. P. Structure of the termini of DNA intermediates in the integration of retroviral DNA: dependence on IN function and terminal DNA sequence. *Cell* **58**, 47–54 (1989).
13. Craigie, R., Fujiwara, T. & Bushman, F. The IN protein of Moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration *in vitro*. *Cell* **62**, 829–837 (1990).
14. Ellison, V. & Brown, P. O. A stable complex between integrase and viral DNA ends mediates human immunodeficiency virus integration *in vitro*. *Proc. Natl. Acad. Sci. USA* **91**, 7316–7320 (1994).
15. DiMaio, F. *et al.* Atomic-accuracy models from 4.5-Å cryo-electron microscopy data with density-guided iterative local refinement. *Nature Methods* **12**, 361–365 (2015).
16. Kudryashov, M. *et al.* Structure of the type VI secretion system contractile sheath. *Cell* **160**, 952–962 (2015).
17. Wang, R. Y. *et al.* *De novo* protein structure determination from near-atomic-resolution cryo-EM maps. *Nature Methods* **12**, 335–338 (2015).
18. van Gent, D. C., Vink, C., Groeneger, A. A. & Plasterk, R. H. Complementation between HIV integrase proteins mutated in different domains. *EMBO J.* **12**, 3261–3267 (1993).
19. Engelman, A., Bushman, F. D. & Craigie, R. Identification of discrete functional domains of HIV-1 integrase and their organization within an active multimeric complex. *EMBO J.* **12**, 3269–3275 (1993).
20. Yang, F. & Roth, M. J. Assembly and catalysis of concerted two-end integration events by Moloney murine leukemia virus integrase. *J. Virol.* **75**, 9561–9570 (2001).
21. Diamond, T. L. & Bushman, F. D. Division of labor within human immunodeficiency virus integrase complexes: determinants of catalysis and target DNA capture. *J. Virol.* **79**, 15376–15387 (2005).
22. Baker, T. A., Mizuuchi, M., Savilahti, H. & Mizuuchi, K. Division of labor among monomers within the Mu transposase tetramer. *Cell* **74**, 723–733 (1993).
23. Jenkins, T. M., Esposito, D., Engelman, A. & Craigie, R. Critical contacts between HIV-1 integrase and viral DNA identified by structure-based analysis and photo-crosslinking. *EMBO J.* **16**, 6849–6859 (1997).
24. Yin, Z., Shi, K., Banerjee, S., Grandgenett, D. P. & Aihara, H. Crystal structure of the Rous sarcoma virus intasome. *Nature* <http://dx.doi.org/10.1038/nature16950> (this issue).
25. Faure, A. *et al.* HIV-1 integrase crosslinked oligomers are active *in vitro*. *Nucleic Acids Res.* **33**, 977–986 (2005).
26. Bera, S., Pandey, K. K., Vora, A. C. & Grandgenett, D. P. Molecular interactions between HIV-1 integrase and the two viral DNA ends within the synaptic complex that mediates concerted integration. *J. Mol. Biol.* **389**, 183–198 (2009).
27. Lee, S. P., Xiao, J., Knutson, J. R., Lewis, M. S. & Han, M. K. Zn²⁺ promotes the self-association of human immunodeficiency virus type-1 integrase *in vitro*. *Biochemistry* **36**, 173–180 (1997).
28. Heuer, T. S. & Brown, P. O. Photo-cross-linking studies suggest a model for the architecture of an active human immunodeficiency virus type 1 integrase-DNA complex. *Biochemistry* **37**, 6667–6678 (1998).
29. Serrao, E., Ballandras-Colas, A., Cherepanov, P., Maertens, G. N. & Engelman, A. N. Key determinants of target DNA recognition by retroviral intasomes. *Retrovirology* **12**, 39 (2015).
30. Engelman, A. & Craigie, R. Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function *in vitro*. *J. Virol.* **66**, 6361–6369 (1992).

Supplementary Information is available in the online version of the paper.

Acknowledgements We acknowledge support from US National Institutes of Health (NIH) grants R01 AI070042 (to A.N.E.), NIH P50 GM103368 and the Leona M. and Harry B. Helmsley Charitable Trust grant number 2012-PG-MED002 (to D.L., both funding sources provided equal support), NIH P50 GM082251 (to P.C.), NIH P30 AI060354 (Harvard University Center for AIDS Research), and US National Science Foundation grants NSF-ACI-1339649 and TG-MCB070039 (to B.D.). B.D. acknowledges support from San Antonio Cancer Institute grant CA054174 for the Center for Analytical Ultracentrifugation of Macromolecular Assemblies at the University of Texas Health Science Center at San Antonio. Molecular graphics and analyses were performed with the USCF Chimera package (supported by NIH P41 GM103331). CryoEM data collection was in part facilitated by the National Resource for Automated Molecular Microscopy (9 P41 GM103310). We thank B. Anderson and J.-C. Ducom at The Scripps Research Institute for help with EM data collection and network infrastructure, J. Fitzpatrick and F. Dwyer for computational support at The Salk Institute, V. Pye for help with X-ray structure refinement and the staff of BM14 (European Synchrotron Radiation Facility, Grenoble, France) and I03 (Diamond Light Source, Oxfordshire, UK) beamlines for assistance with data collection.

Author Contributions A.B.-C. and A.N.E. discovered how to assemble MMTV intasomes; A.B.-C. and T.G.D. expressed and purified MMTV IN proteins for biochemical analysis; A.B.-C. assembled intasomes, characterized their biochemistry, supplied them for cryo-EM and centrifugation analyses, and performed IN activity assays; M.B. and D.L. performed EM work, collected cryo-EM data and determined the structure; D.L. modelled the intasome structure; B.D. collected and analysed the sedimentation velocity data; N.J.C. and P.C. expressed and purified IN_{CCD}, IN_{NTD-CCD} and IN_{CTD-212-266} constructs, established crystallization conditions and determined these structures.

Author Information Coordinates of cryo-EM density maps for the full and core intasome datasets have been deposited in the Electron Microscopy Data Bank under accession numbers EMD-6440 and EMD-6441, respectively. X-ray diffraction data and the resulting IN_{CCD}, IN_{NTD-CCD} and IN_{CTD} structures have been deposited in the Protein Data Bank (PDB) under accession numbers 5CZ1, 5CZ2 and 5D7U, respectively. The core intasome structure has been deposited in the Protein Data Bank under accession number 3JCA. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.L. (dlyumkis@salk.edu) or A.N.E. (alan_engelman@dfci.harvard.edu).

METHODS

Statistical methods were not used to predetermine sample sizes. Experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

DNA constructs. Full-length (FL) MMTV IN³¹ and IN_{NTD} (IN₂₁₂₋₂₆₆ and IN₂₁₂₋₃₁₉) expression constructs provided N-terminal His₆ tags followed by human rhinovirus (HRV) 3C protease cleavage sites. The IN_{NTD-CCD} expression construct was made by introducing a stop codon after the TCA that encodes for IN residue Ser212. IN_{K165E} and IN_{R240E} expression constructs were made by PCR-directed mutagenesis. DNA fragments corresponding to IN₅₁₋₂₁₂ (IN_{CCD}) and IN₅₁₋₃₁₉ (IN_{CCD-CTD}) were amplified by PCR and subcloned into expression vector pET-20b (Novagen); these proteins harboured cleavable C-terminal His₆ tags. The sequences of all PCR amplified regions of plasmid DNAs were verified by sequencing.

Protein expression and purification for intasome and IN activity assays. FL INs, IN_{CCD-CTD} and IN_{CTD/212-319} were expressed in *Escherichia coli* strain PC2 (ref. 32) in LB broth (supplemented with 50 μM ZnCl₂ for FL INs) by induction with 0.4 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) (1 mM IPTG for IN_{CCD-CTD}) at 30 °C (37 °C for IN_{CCD-CTD} and IN_{CTD}) for 4 h. Bacteria pellets were resuspended in 20 mM HEPES, pH 7.6, 1 M NaCl, 5 mM 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS), complete EDTA-free protease inhibitor (Roche). After sonication for 5 min at 50 mA, cell lysates were centrifuged at 45,000 g for 1 h. The supernatant, supplemented with 5 mM imidazole, was filtered through a 0.45 μm filter and purified using a Ni²⁺-charged HisTrap 5 ml column (GE Healthcare) equilibrated with 20 mM HEPES, pH 7.6, 1 M NaCl, 5 mM CHAPS, 15 mM imidazole. Proteins were eluted by a linear gradient of imidazole (15–500 mM) containing a step wash at 65 mM imidazole using the ÄKTA purifier system (GE Healthcare; for IN_{CCD-CTD}, a second step wash was done at 115 mM imidazole). IN-containing fractions were diluted 1:5 with 20 mM HEPES, pH 7.6, 5 mM CHAPS, 2 mM dithiothreitol (DTT) and immediately loaded on a Heparin HiTrap 5 ml column equilibrated with 20 mM HEPES, pH 7.6, 200 mM NaCl, 5 mM CHAPS, 2 mM DTT. Proteins were eluted by a linear NaCl gradient from 200 mM to 2 M (IN_{CTD} was isolated in the column flow through). IN-containing fractions were pooled and cleaved with HRV 3C protease (GE Healthcare) overnight at 4 °C to remove the His₆ tag. In lieu of purification by Heparin HiTrap, IN_{CCD-CTD} was dialysed against 20 mM HEPES, pH 7.6, 1 M NaCl, 5 mM CHAPS, 2 mM DTT, 2 mM EDTA at 4 °C for 2 h, cleaved with HRV 3C protease overnight at 4 °C, followed by dialysis against 20 mM HEPES, pH 7.6, 1 M NaCl, 5 mM CHAPS, 2 mM DTT, 0.5 mM EDTA (SEC1 buffer). Cleaved proteins were purified by SEC using a Superdex 200 10/300 column (GE Healthcare) equilibrated with SEC1 buffer. Purified INs were concentrated by ultracentrifugation using 10-kDa molecular mass cutoff Millipore concentrators and dialysed overnight against SEC1 buffer supplemented to contain 10% glycerol. Protein concentration was determined by spectrophotometry, and aliquots flash-frozen in liquid N₂ were stored at –80 °C. Purified INs were analysed by SEC using a Superdex 3.2/300 column equilibrated with SEC1 buffer; protein standards were from Bio-Rad.

MMTV intasome assembly. Intasomes were assembled by mixing 128 μM MMTV IN with 38 μM 22 bp preprocessed vDNA (5'-CAGGTCGGCCGACTGCGGCA/5'-AATGCCGAGTCGGCCGACCTG) in 20 mM HEPES, pH 7.6, 600 mM NaCl, 2 mM DTT, before dialysis for 16 h at 4 °C against 25 mM Tris-HCl, pH 7.4, 80 mM NaCl, 2 mM DTT, 25 μM ZnCl₂, 10 mM CaCl₂. The resulting milky white precipitate was dissolved by adding NaCl to the final concentration of 250 mM, followed by incubation on ice for 1 h. After centrifugation for 10 min at 20,000 g at 4 °C, soluble intasomes were purified by SEC using Superdex 200 10/300 equilibrated with 25 mM Tris-HCl, pH 7.4, 200 mM NaCl, 2 mM DTT, 25 μM ZnCl₂, 10 mM CaCl₂ (SEC2 buffer). Intasome-containing fractions, which eluted around 10.5 ml, were concentrated by ultracentrifugation using 10-kDa cut off concentrators.

In vitro integration assays. Strand transfer assays were performed as described previously³¹. Briefly, 1 μM intasome or 1 μM MMTV IN plus 0.5 μM vDNA were mixed with 300 ng pGEM-3 tDNA in 40 μl of 20 mM HEPES, pH 7.4, 60 mM NaCl, 5 mM MgCl₂, 4 μM ZnSO₄, 10 mM DTT. Reactions incubated for 1 h at 37 °C were terminated by adding 25 mM EDTA–0.5% SDS. DNA products deproteinized by digestion with proteinase K and precipitated with ethanol were analysed by electrophoresis through 1.5% agarose gels and visualized by staining with ethidium bromide. Raltegravir, which was used at the final concentration of 100 μM, was obtained from Selleck Chemicals. Proteins were premixed on ice before addition to reactions for biochemical complementation assays. Concerted integration products were measured by band intensity quantification relative to IN_{WT} product formation, which was set to 100% using Molecular Imager Gel Doc TM XR+ System with Image Lab software (BioRad); the background across eight gel images corresponded to 1.26% ± 0.47% of IN_{WT} function.

Concerted integration reaction products were cloned and sequenced essentially as previously described³². Briefly, DNA excised from agarose gels was repaired using Phi29 DNA polymerase (New England Biolabs) and ligated to a

PCR-amplified kanamycin resistance cassette. Plasmids recovered after transformation of ligation mixtures into *E. coli* were sequenced using primers that annealed to the ends of the cassette DNA.

Analytical ultracentrifugation. We analysed sedimentation velocity at 20 °C in a Beckman Optima XL-I analytical ultracentrifuge using an An60Ti rotor and standard two-channel Epon Centerpieces (Beckman-Coulter). Samples were prepared in 20 mM phosphate buffer, pH 7.5, 150 mM NaCl at two loading concentrations, absorbance ($A_{280\text{ nm}}$) values of 0.3 and 0.9 for MMTV IN and the intasome, and $A_{280\text{ nm}}$ values of 0.18 and 0.53 for vDNA, to exclude potential mass action oligomerization. IN and vDNA were spun simultaneously at 35,000 r.p.m. for 22 h while the intasome was spun at 27,000 r.p.m. for 12 h; the different rotor speeds were based on the predicted masses of the different macromolecules.

Data were analysed using UltraScan-III version 2.2, release 2000 (ref. 33). Hydrodynamic corrections for buffer density and viscosity were estimated with UltraScan to be 1.041 g ml⁻¹ and 1.101 centipoise, respectively. The partial specific volume of IN (0.728 ml g⁻¹) was estimated by UltraScan from its protein sequence using a method analogous to the methods outlined in ref. 34. Sedimentation velocity data were analysed as described³⁵. Optimization was performed by two-dimensional spectrum analysis³⁶ with simultaneous removal of time-invariant and radially-invariant noise contributions³⁷. Two-dimensional spectrum analysis solutions, which are subjected to parsimonious regularization by genetic algorithm analysis³⁸, were further refined using Monte Carlo analysis to determine confidence limits for the determined parameters³⁹. Calculations were performed on the Lonestar cluster at the Texas Advanced Computing Center at the University of Texas at Austin.

Protein expression and X-ray crystallography. MMTV IN_{CCD}, IN_{NTD-CCD} and IN_{CTD} fragments spanning MMTV IN residues 51–212, 1–212 and 212–266, respectively, were expressed in BL21(DE3)-CodonPlus cells (Stratagene) in LB medium (supplemented with 50 μM ZnCl₂ for IN_{NTD-CCD}) by induction with 0.01% (w/v) IPTG. Bacteria were lysed by sonication in 0.5 M NaCl, 50 mM Tris-HCl, pH 7.4, and the proteins were isolated by absorption to Ni-nitrilotriacetic acid agarose (Qiagen). After digestion with HRV 3C protease to release His₆ tags, the proteins were further purified by ion exchange and SEC.

Crystals were grown by vapour diffusion in hanging drops by mixing 1 μl protein (6–10 mg ml⁻¹ in 200 mM NaCl, 2 mM DTT, 25 mM Tris-HCl, pH 7.5) and 1 μl reservoir solution, which contained 12.5% PEG-3350, 0.15 M ammonium citrate, pH 6.5 (IN_{CCD}), 19% PEG-3350, 0.2 M MgCl₂, 5% (v/w) 1-butyl-3-methylimidazolium dicyanamide (IN_{NTD-CCD}) or 19% isopropanol, 50 mM ammonium acetate, 0.1 M HEPES-NaOH, pH 7.5 (IN_{CTD}). Crystals, cryoprotected with 25% glycerol (IN_{CCD}, IN_{NTD-CCD}) or 30% PEG-400 (IN_{CTD}), were frozen by immersion in liquid nitrogen. Diffraction data for the IN_{CCD} were collected using a charge-coupled device detector at beamline BM14 (European Synchrotron Radiation Facility) whereas IN_{CTD} and IN_{NTD-CCD} crystals were analysed at beamline I03 (Diamond Light Source) equipped with a PILATUS direct detector. The data, integrated with XDS⁴⁰, were scaled with Aimless⁴¹. The structures, which were each derived from a single crystal, were solved by molecular replacement in Phaser⁴² using search models generated from PDB entries 1ASV (CCD)⁴³, 3F9K (NTD)¹⁰ and 1EX4 (CTD)⁴⁴. The models were rebuilt using ARP/wARP⁴⁵ and/or manually in Coot⁴⁶ and refined in Phenix⁴⁷ and/or Refmac⁴⁸. Pseudo-merohedral twin law ($-h, -k, l$) was accounted for during refinement of the IN_{NTD-CCD} structure. Final models, validated with MolProbity⁴⁹, had at least 96.9% of residues in the favoured regions and none in the disallowed regions of the Ramachandran plot. Detailed X-ray data collection and refinement statistics are given in Extended Data Table 1.

Cryo-EM data acquisition. Sample containing MMTV intasomes in SEC2 buffer supplemented to contain 0.05% NP-40 was applied onto freshly plasma treated (6 s, Gatan Solarus plasma cleaner) holey carbon C-flat grids (CF-1.2/1.3-4C, Protochips), adsorbed for 30 s and then plunged into liquid ethane using a manual cryo-plunger in an ambient environment of 4 °C.

Data were acquired over three separate sessions using Leginon software⁵⁰ installed on an FEI Titan Krios electron microscope operating at 300 kV, with a dose of 40 electrons per pixel per square ångström at a rate of ~6.9 electrons per pixel per second and an estimated underfocus ranging from 1 to 4 μm (centred at 2.6 ± 0.6 μm). The dose was fractionated over 50 raw frames collected over a 10-s exposure time (200 ms per frame) on a Gatan K2 Summit direct detection device, with each frame receiving a dose of ~6.5 electrons per pixel per second. Two thousand seven hundred and fourteen movies were collected and recorded at a nominal magnification of 22,500, corresponding to a pixel size of 1.31 Å at the specimen level. The individual frames were gain corrected, aligned and summed using a graphic processing unit-enabled whole-frame alignment program as previously described^{51,52}, and exposure filtered⁵³ according to a dose rate of 6.9 electrons per pixel per second. See Extended Data Table 2 for additional details on cryo-EM data collection.

Cryo-EM image analysis. Pre-processing operations before the refinement of the final models were performed using the Appion package⁵⁴ and were conceptually

identical to those previously described⁵². Briefly, single intasome particles (244,315) were selected from the aligned and summed micrographs, from which 147,850 were used to create an initial raw particle stack after removing regions of the micrographs containing carbon and large areas of aggregation. Two-dimensional alignments and classifications were performed using the CL2D⁵⁵ and Relion⁵⁶ algorithms (Extended Data Fig. 1c), and an initial model was generated directly from the class averages using OptiMod⁵⁷ (Extended Data Fig. 1d). After iterative rounds of two-dimensional alignment and classification, 77,365 particles remained for three-dimensional refinement and classification. Three-dimensional refinements and classifications were initially performed within Relion⁵⁶, after which the parameters were converted for use in Frealign⁵⁸. The final map was refined in Frealign.

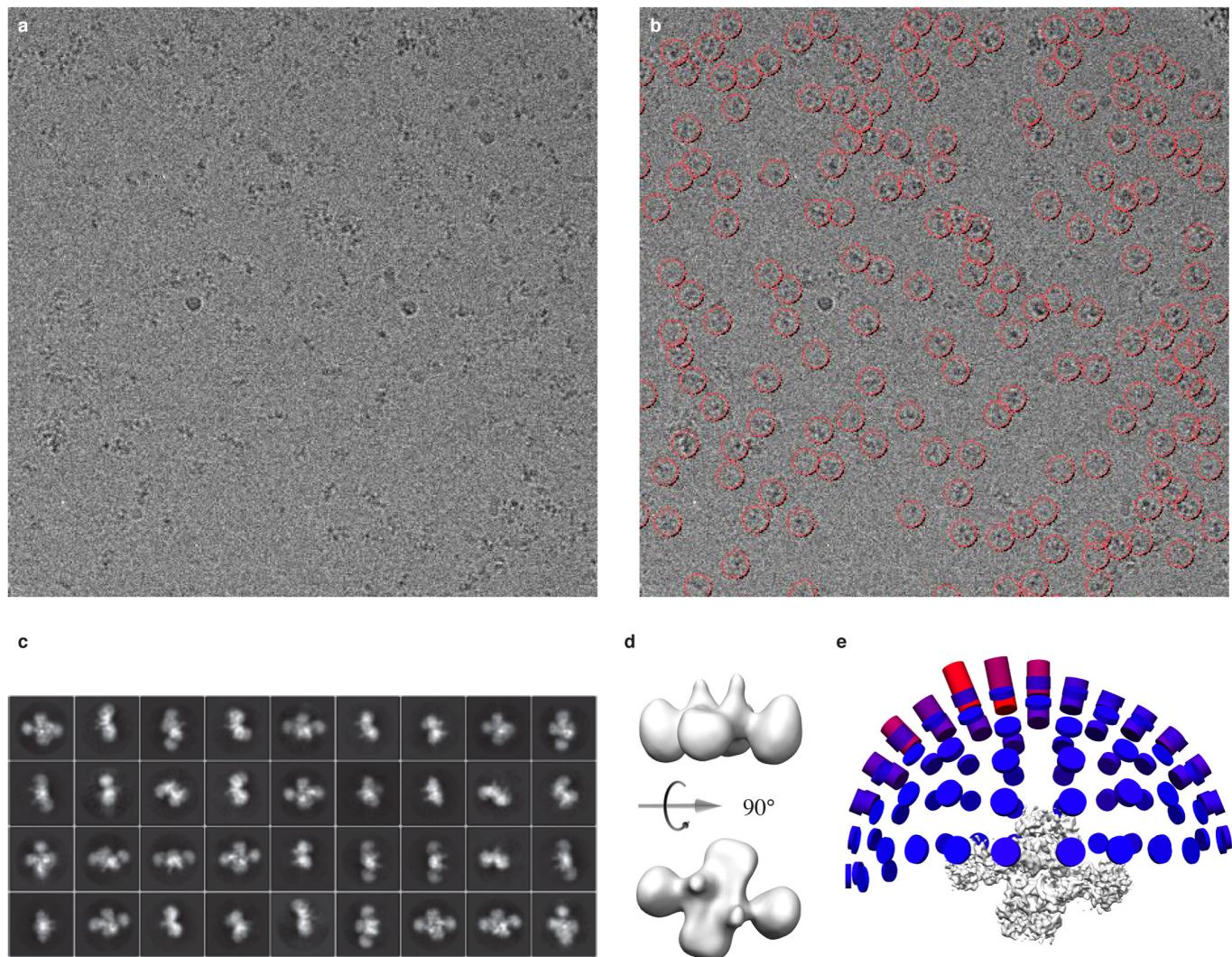
Several conformational states of the intasome were observed after three-dimensional classification in both Relion and Frealign⁵⁹. Whereas one of the resulting maps yielded the stable intasome structure from 41,475 particles (Fig. 2a, Extended Data Fig. 2c and Extended Data Table 2), all other maps (one of which is shown in Extended Data Fig. 3b) displayed mobility in the flanking regions, which did not resolve by further classifying the data. To improve the resolution of the core region, we ran Relion and recovered four models in the classification. For each of the resulting maps, the flanking regions were segmented and treated with a soft-edged mask that adopted the shape of the remaining density. Subsequently, for each raw particle, the flanking region from the respective conformational state to which that particle belonged was computationally subtracted from the raw particle image. The contrast transfer function was included in the computational subtraction process. In this manner, data sets lacking most of the flanking INs were created. Refinement of the core intasome data set was then conducted using the likelihood-based approach in Frealign⁵⁹, effectively a focused classification of the core region. The best class was resolved to ~ 4 Å resolution in the most homogeneous regions using 30,307 particles (Extended Data Fig. 2d and Extended Data Table 2). Although slight ghost images remained for the flanking regions within certain particles, they did not dramatically affect the refinement; the use of a tighter mask facilitated the recovery of higher-resolution information.

Assembly of the atomic model. Models of the core intasome and the full octamer structures were built and refined in a stepwise manner using Rosetta¹⁵ starting with rigid-body fitted X-ray structures of individual domains as input. Rosetta protocols were used for all parts of the modelling⁶⁰. To optimally fit X-ray models into the EM density, we first independently refined each individual domain (NTD, CCD and CTD) using multiple-input starting seeds. CCD₁ and CCD₂ were each seeded with six starting X-ray models: independent CCD monomers from chains A–D of the IN_{CCD} structure and monomers A–B of the CCD portions of the IN_{NTD–CCD} structures. CTDs 1, 2, 5 and 6 were seeded with subunits A and B of the IN_{CTD} X-ray model. Likewise, for NTD₁ and NTD₃, the two different NTDs of the IN_{NTD–CCD} X-ray structure were used as input seeds. All models were refined against the core intasome structure resolved to ~ 4 –5 Å resolution (Extended Data Fig. 2d). At least 2,000 models were generated from each and the lowest-energy model was selected for moving forward. Modelling quality was assessed by energy scores, structural similarity of the top scoring models and visual inspection (Extended Data Fig. 6a). We next proceeded to independently model IN₁, IN₂, IN₅ and IN₆, thereby filling in the linker regions between individual domains using seven-amino-acid oligopeptides from the PDB¹⁵. This enabled *de novo* modelling for linker residues 45–54 between NTD₁–CCD₁ and residues 211–213 between CCD₁–CTD₁ and CCD₂–CTD₂ (some residues, as well as outlier linker regions, were not modelled owing to disorder; Extended Data Fig. 6b, c); modelling was facilitated by the presence of ‘bumps’ within the density that corresponded to bulky amino-acid side chains, in particular within NTD₁–CCD₁, which is located in the best-resolved region of the structure (Extended Data Fig. 2d). IN₁ and IN₂ were each seeded with combinations of the best models arising from refinement of individual domains and were subsequently refined against the core intasome density map. Two thousand models were again generated for each, and the best were selected to move forward. This set of procedures produced FL models for IN₁ and IN₂ and models for CTD₅ and CTD₆ fitted to the EM protein density. MMTV DNA was modelled on the basis of the X-ray structure of the PFV intasome (PDB accession number 3L2Q). This model was rigid-body docked into the EM density and then relaxed with Rosetta. The complete intasome model was iteratively relaxed with Rosetta and then adjusted manually using Coot⁴⁶. Several iterative rounds of refinement and inspection were performed using MolProbity⁴⁹ at the end of each round until a consensus model was obtained (Extended Data Fig. 6c, d and Extended Data Table 2).

IN linker regions. Linker lengths for Extended Data Fig. 8b were assessed by aligning published³⁰ or in-house generated IN sequence alignments against alignments based on known domain structures² (Extended Data Fig. 4a). The following sequences were included: gammaretroviruses: Moloney murine leukaemia virus (GenBank accession number J02255.1), reticuloendotheliosis virus strain A (DQ237900.1), feline leukaemia virus (NC_001940.1); epsilonretroviruses: walleye

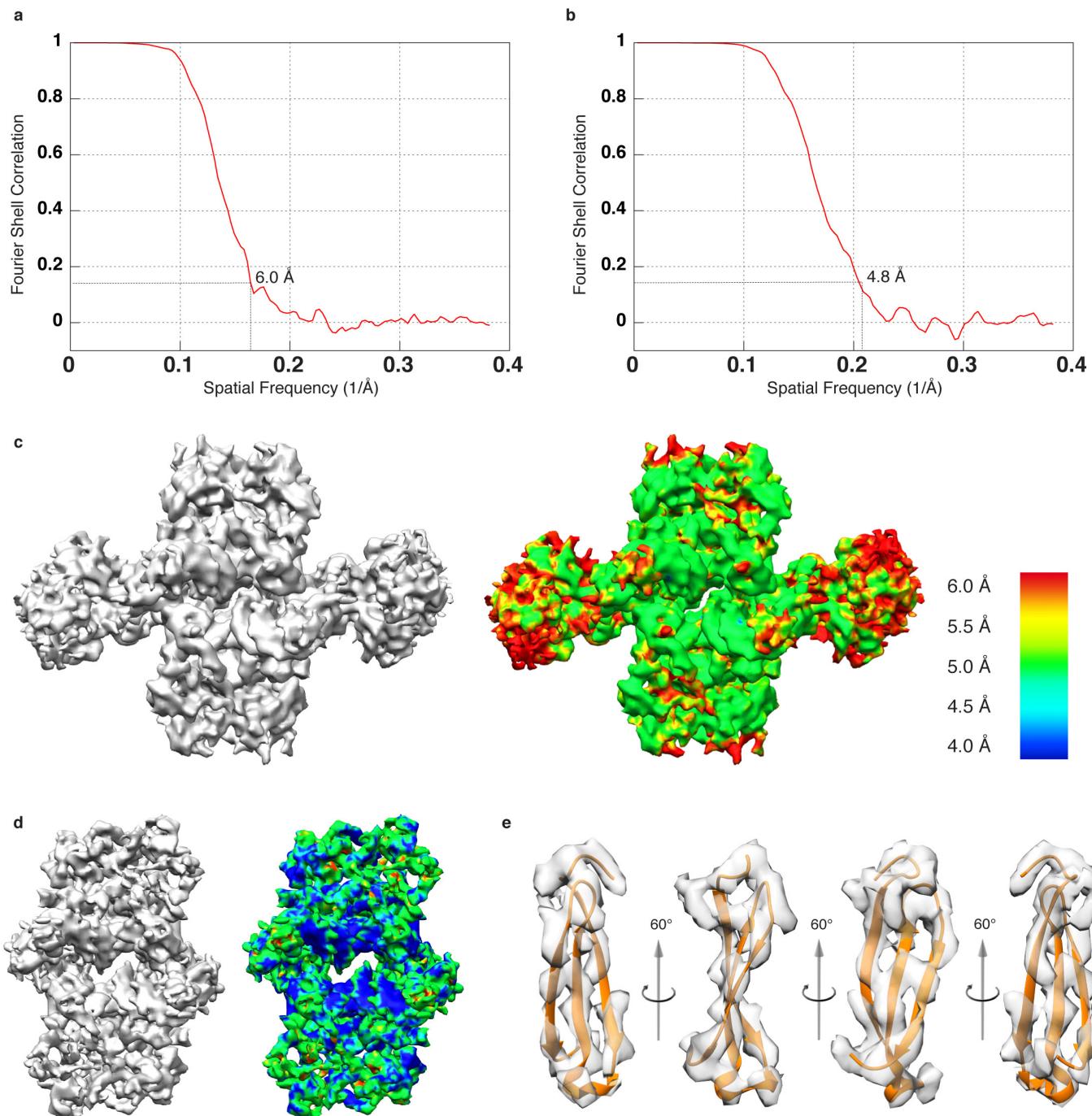
dermal sarcoma virus (NC_001867.1), walleye epidermal hyperplasia virus types 1 and 2 (AF133051.1 and AF133051.2, respectively); spumaviruses: PFV (U21247.1), macaque simian foamy virus (NC_010819.1), spider monkey foamy virus (EU010385.1); lentiviruses: HIV-1 strain NL4-3 (U26942.1), HIV-2 strain ROD (X05291.1), simian immunodeficiency virus strain agm.tan-1 (U58991.1), equine infectious anaemia virus (M16575.1), feline immunodeficiency virus (M25381.1), caprine arthritis encephalitis virus (M33677.1), bovine immunodeficiency virus (NC_001413.1); deltaretroviruses: bovine leukaemia virus (K02120.1), human T-cell lymphotropic virus types 1 and 2 (NC_001436.1 and NC_001488.1, respectively); betaretroviruses: MMTV (NC_001503.1), Mason Pfizer monkey virus (NC_001550.1), Jaagsiekte sheep retrovirus (NC_001494.1); alpharetroviruses: Rous sarcoma virus (J02342.1), lymphoproliferative disease virus (KC802224.1).

- Ballandras-Colas, A., Naraharisetti, H., Li, X., Serrao, E. & Engelman, A. Biochemical characterization of novel retroviral integrase proteins. *PLoS ONE* **8**, e76638 (2013).
- Cherepanov, P. LEDGF/p75 interacts with divergent lentiviral integrases and modulates their enzymatic activity *in vitro*. *Nucleic Acids Res.* **35**, 113–124 (2007).
- Demeler, B. *UltraScan-III* version 2.2: a comprehensive data analysis software package for analytical ultracentrifugation experiments <<http://www.ultrascan3.uthscsa.edu/>> (2014).
- Laue, T. M., Shah, B. D., Ridgeway, T. M. & Pelletier, S. L. in *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (eds Harding, S. E., Rowe, A. J. & Horton, J. C.) 90–125 (Royal Society of Chemistry, 1992).
- Demeler, B. in *Current Protocols in Protein Science* (eds Coligan, J. E. *et al.*) Ch. 7, Unit 7.13, 7.13.1–7.13.24 (Wiley, 2010).
- Brookes, E., Cao, W. & Demeler, B. A two-dimensional spectrum analysis for sedimentation velocity experiments of mixtures with heterogeneity in molecular weight and shape. *Eur. Biophys. J.* **39**, 405–414 (2010).
- Schuck, P. & Demeler, B. Direct sedimentation analysis of interference optical data in analytical ultracentrifugation. *Biophys. J.* **76**, 2288–2296 (1999).
- Brookes, E. H. & Demeler, B. in *Proceedings of the 9th Annual Conference on Genetic and Evolutionary Computation* 361–368 (Association for Computing Machinery, 2007).
- Demeler, B. & Brookes, E. Monte Carlo analysis of sedimentation experiments. *Colloid Polym. Sci.* **286**, 129–137 (2008).
- Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
- Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D* **69**, 1204–1214 (2013).
- McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
- Bujacz, G. *et al.* High-resolution structure of the catalytic domain of avian sarcoma virus integrase. *J. Mol. Biol.* **253**, 333–346 (1995).
- Chen, J. C.-H. *et al.* Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding. *Proc. Natl. Acad. Sci. USA* **97**, 8233–8238 (2000).
- Morris, R. J., Perrakis, A. & Lamzin, V. S. ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol.* **374**, 229–244 (2003).
- Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
- Adams, P. D. *et al.* The Phenix software for automated determination of macromolecular structures. *Methods* **55**, 94–106 (2011).
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
- Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
- Suloway, C. *et al.* Automated molecular microscopy: the new Legion system. *J. Struct. Biol.* **151**, 41–60 (2005).
- Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nature Methods* **10**, 584–590 (2013).
- Lyumkis, D. *et al.* Cryo-EM structure of a fully glycosylated soluble cleaved HIV-1 envelope trimer. *Science* **342**, 1484–1490 (2013).
- Grant, T. & Grigorieff, N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife* **4**, e06980 (2015).
- Lander, G. C. *et al.* Appion: an integrated, database-driven pipeline to facilitate EM image processing. *J. Struct. Biol.* **166**, 95–102 (2009).
- Sorzano, C. O. *et al.* A clustering approach to multireference alignment of single-particle projections in electron microscopy. *J. Struct. Biol.* **171**, 197–206 (2010).
- Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
- Lyumkis, D., Vinterbo, S., Potter, C. S. & Carragher, B. Optimod – an automated approach for constructing and optimizing initial models for single-particle electron microscopy. *J. Struct. Biol.* **184**, 417–426 (2013).
- Grigorieff, N. FREALIGN: high-resolution refinement of single particle structures. *J. Struct. Biol.* **157**, 117–125 (2007).
- Lyumkis, D., Brilot, A. F., Theobald, D. L. & Grigorieff, N. Likelihood-based classification of cryo-EM images using FREALIGN. *J. Struct. Biol.* **183**, 377–388 (2013).
- DiMaio, F., Zhang, J., Chiu, W. & Baker, D. Cryo-EM model validation using independent map reconstructions. *Protein Sci.* **22**, 865–868 (2013).
- Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).


Extended Data Figure 1 | Cryo-EM data and refinement.

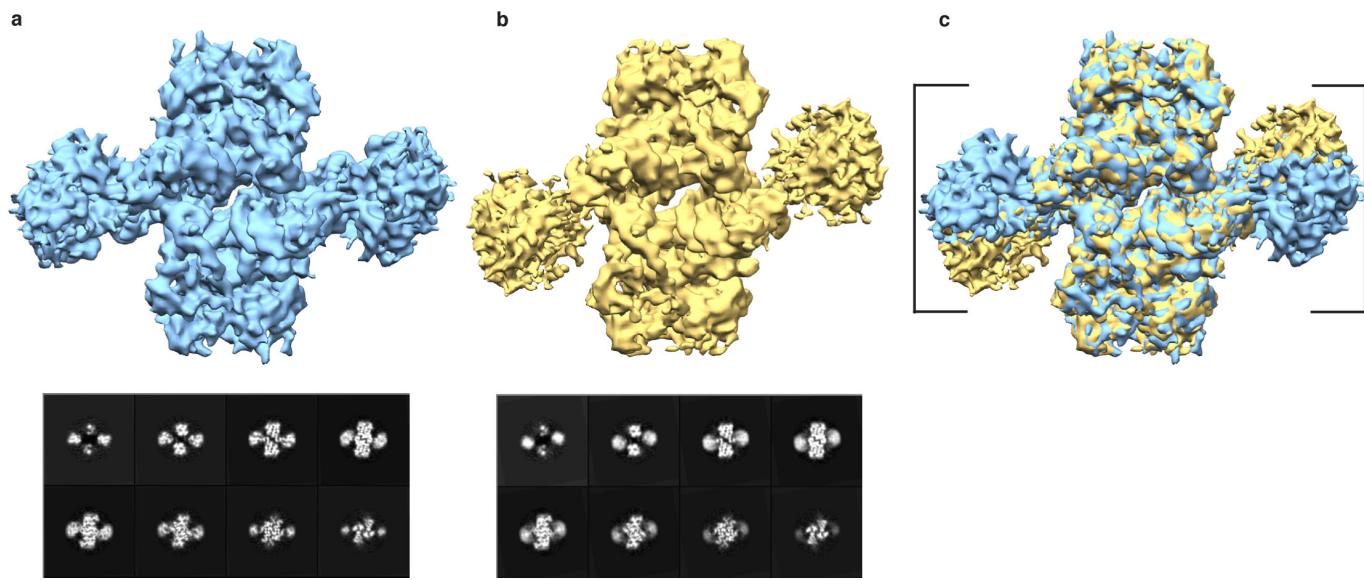
a, Representative cryo-electron micrograph of MMTV intasomes, taken at $2.7\text{ }\mu\text{m}$ underfocus. **b**, Same as in **a**, marked to show selected particles. **c**, Two-dimensional class averages calculated using Relion⁵⁶. **d**, Initial

model from the class averages calculated using OptiMod⁵⁷. **e**, Refined reconstruction from the full data set, with an Euler angle distribution plot showing the relative orientations of the particles.



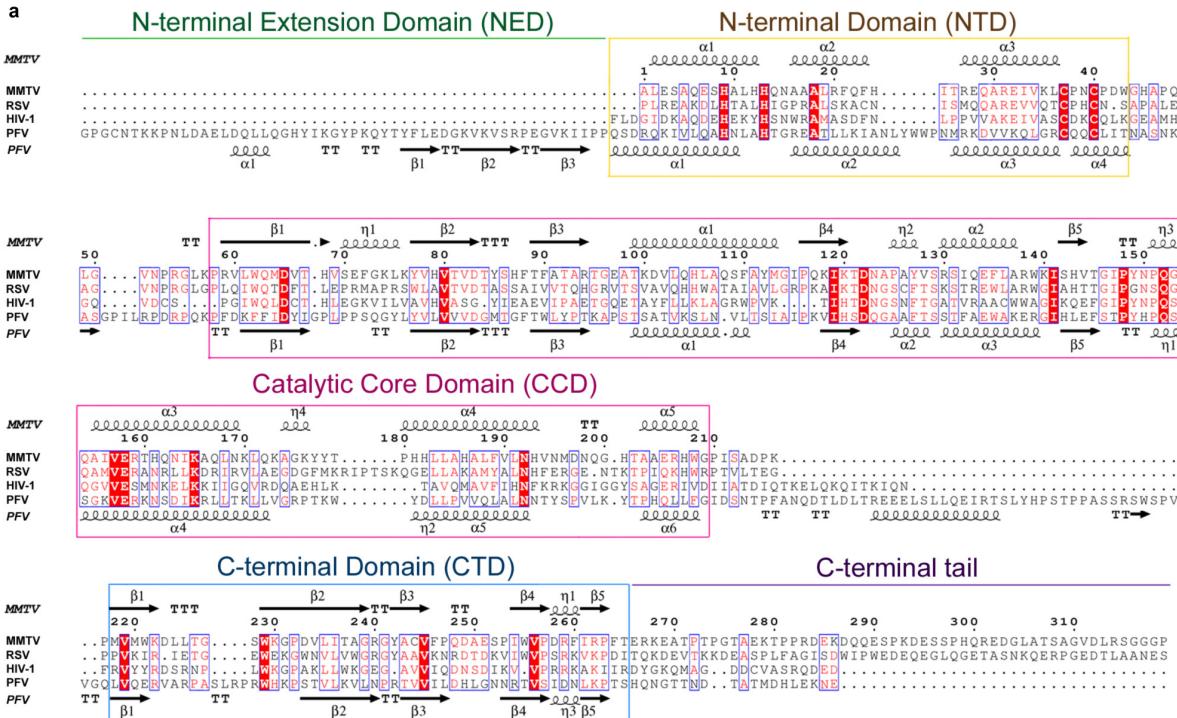
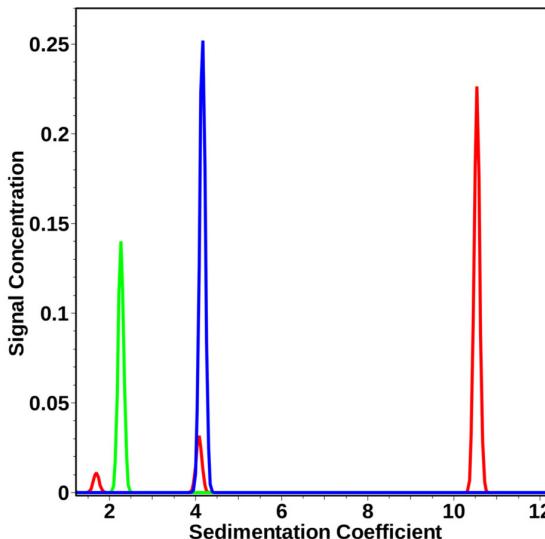
Extended Data Figure 2 | Cryo-EM resolution analysis of reconstructed intasome maps. **a**, Fourier shell correlation curve corresponding to the refined map generated from the full intasome data set. **b**, Fourier shell correlation curve corresponding to the refined map generated from the core intasome data set with the NTDs, CCDs and interdomain linker regions of the flanking IN dimers computationally subtracted. Average global resolutions in **a** and **b** are indicated. **c**, Refined map generated from the full data set (left) displayed side-by-side with the same map coloured

for local resolution (right). **d**, Refined map generated from the core intasome data set (left) displayed side-by-side with the same map coloured for local resolution (right) using the colouring scheme in **c**. **e**, Rotational snapshots of segmented density of CCD₁ with the fit of the refined model (see Extended Data Fig. 6) highlighting structural features evident at ~4–5 Å resolution. Partial separation of β -strands, which is typically evident at or beyond 4.5 Å resolution, is apparent.



Extended Data Figure 3 | Structural heterogeneity of the MMTV intasome. **a**, Stable structural conformation of the MMTV intasome after three-dimensional classification of the data. Slices from the density map are displayed below. **b**, One of several conformations of MMTV intasome refinement after three-dimensional classification of the data. Slices from

the density map are displayed below. Multiple fuzzy regions in the flanking INs are apparent in **b**, which are indicative of remaining heterogeneity within the data and/or continuous structural mobility of the region. **c**, Overlay of the two reconstructed maps, highlighting the extent of mobility within the flanking regions (brackets).

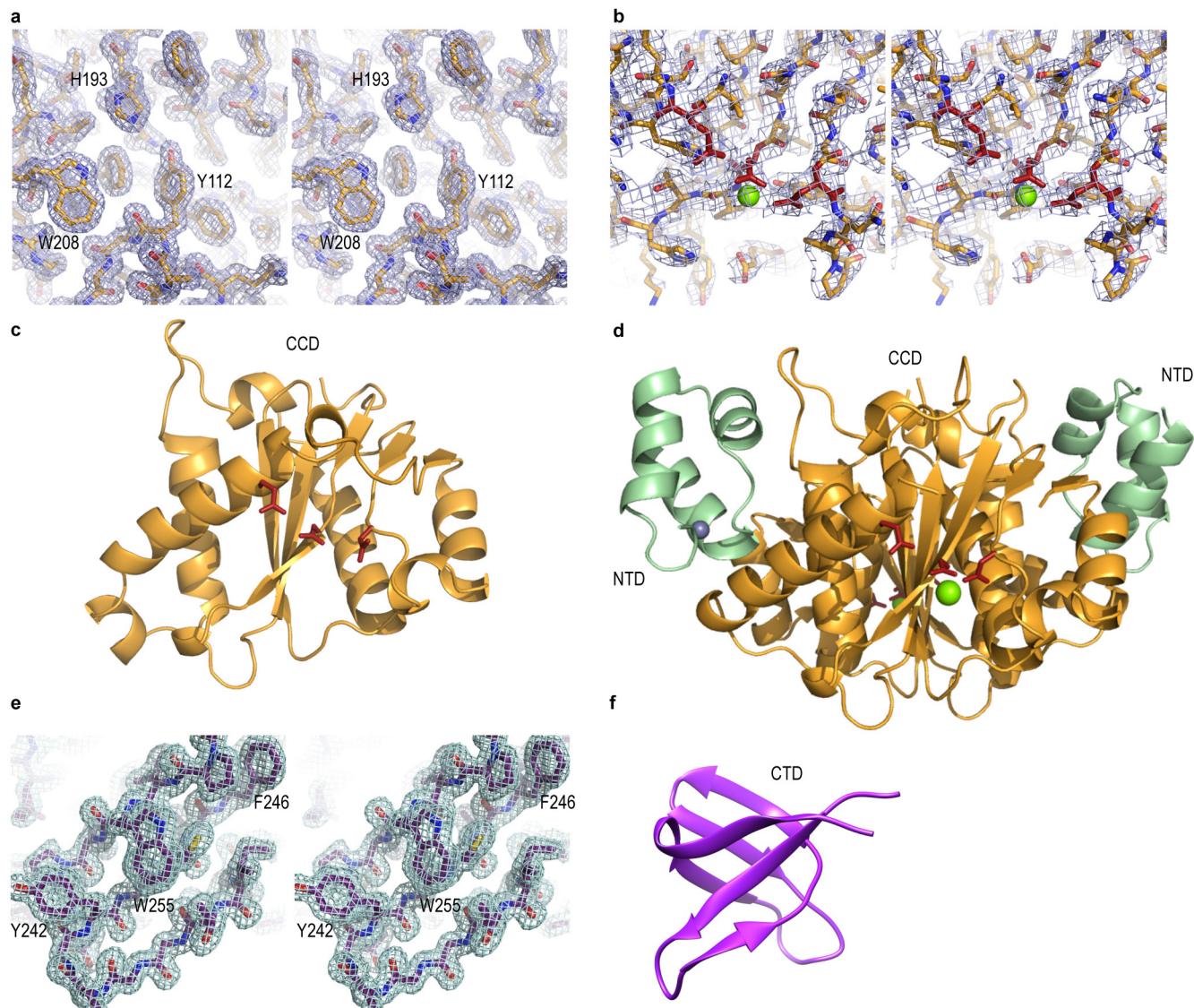
a**b**

Species	Molar mass (kDa) (measured)	Molar mass (kDa) (theoretical)	s ($\times 10^{-13}$ sec)
vDNA	11.9	13.2	2.27
IN	40.9	35.9	3.89
intasome	302.1*	313.6	10.53

* assuming a partial specific volume weight-average of 0.713 ml/g for an 8:2 protein:DNA complex.

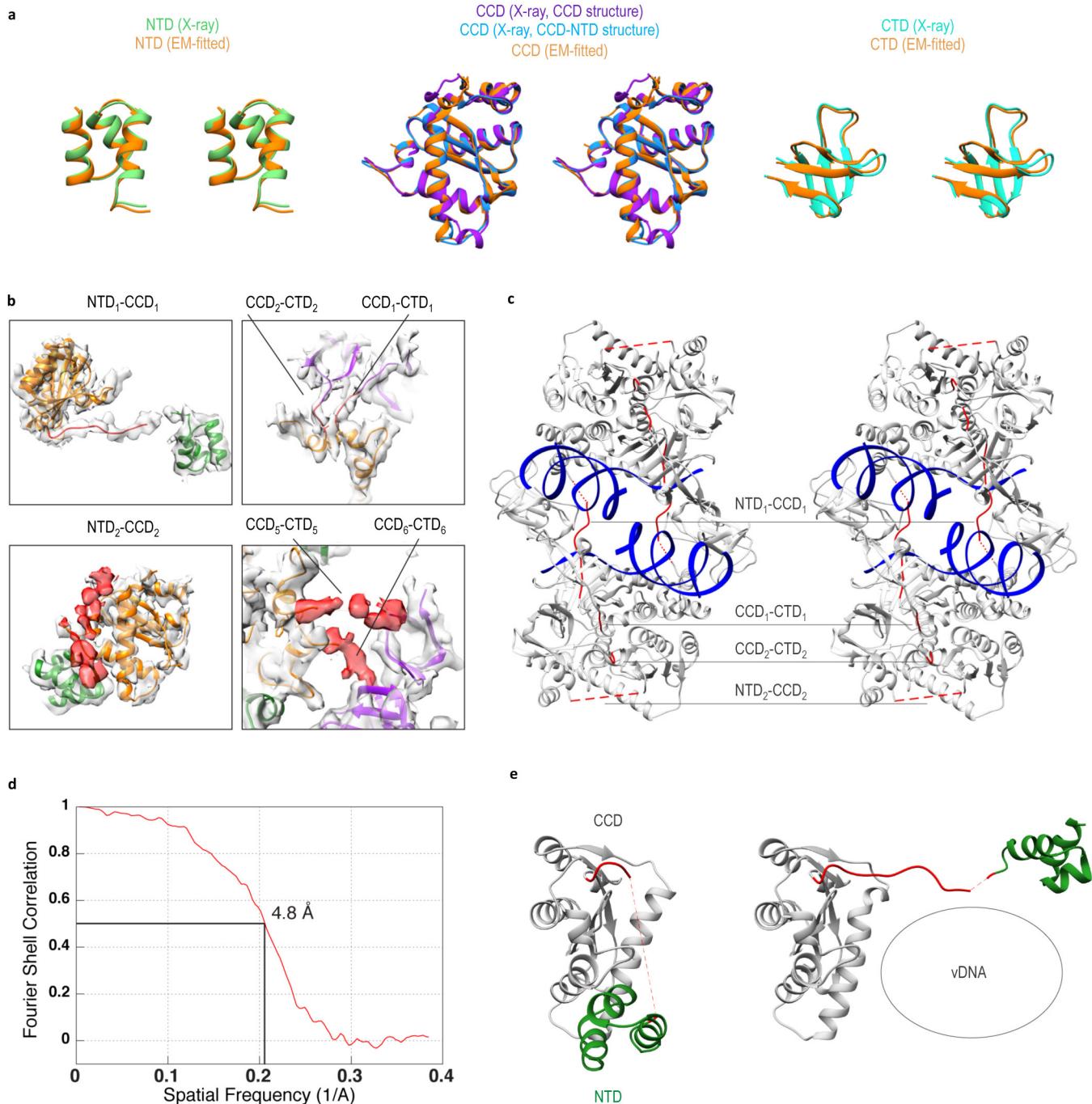
Extended Data Figure 4 | MMTV IN domains and intasome sedimentation coefficient distribution. **a**, Primary IN sequence alignment with boxes denoting canonical IN structural domains. The N-terminal extension domain occurs in spuma-, gamma- and epsilonretroviral IN proteins. Identical residues between MMTV, Rous sarcoma virus, HIV-1 and PFV INs are highlighted by red background; residues that are minimally conserved in three of the sequences are in red. PFV IN secondary structure elements are from PDB accession number 3L2Q; MMTV elements are from the IN_{NTD-CCD} and IN_{CTD} crystal structures described here (PDB accession numbers 5CZ2 and 5D7U, respectively). Symbols α , β , η , TT and TTT represent α -helix, β -strand,

β_{10} -helix, α -turn and β -turn, respectively. Figure generated with ESPript 3.0 (ref. 61). **b**, Monte Carlo analysis of sedimentation velocity data for the higher loading concentrations of vDNA (green), MMTV IN (blue) and intasome (red). A clear shift to a discrete species at 10.5 s is observed for the intasome, with minor IN and vDNA populations evident. Different centrifugation parameters for IN and vDNA versus intasomes (see Methods) probably attributed to the minor variations in sedimentation coefficient between major and minor IN and vDNA species. Measured sedimentation coefficients and calculated molar masses compared with theoretical molar masses are shown beneath the graph.



Extended Data Figure 5 | MMTV IN domain crystal structures. **a**, Stereo view of the final $2F_o - F_c$ density map of the IN_{CCD} crystal structure with blue mesh contoured at 1σ . Amino-acid side chains are readily evident at the 1.7 \AA resolution. **b**, Stereo view of the final $2F_o - F_c$ density map of the 2.7 \AA resolution IN_{NTD-CCD} crystal structure with blue mesh contoured at 1σ . The map is centred on the DDE catalytic triad (red sticks); green spheres, Mg^{2+} ions. **c**, Cartoon representation of the IN_{CCD} monomer (one of four in the crystallographic asymmetric unit) coloured in gold.

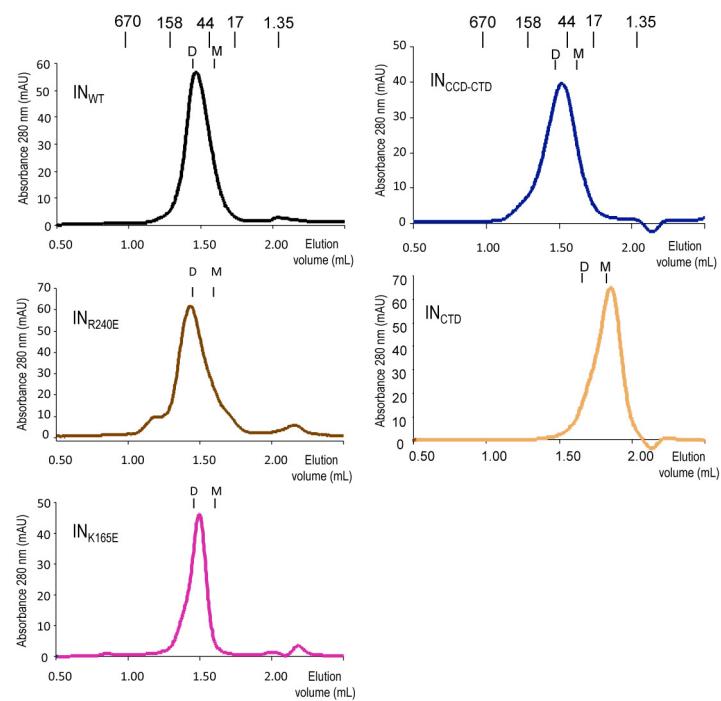
Active site residues are shown as red sticks. **d**, Cartoon representation of the IN_{NTD-CCD} dimer structure (one of three in the asymmetric unit). The NTD and CCD are coloured green and gold, respectively. Red sticks, active site residues; grey and green spheres, Zn^{2+} and Mg^{2+} ions, respectively. **e**, Stereo view of the final $2F_o - F_c$ density map of the 1.5 \AA resolution IN_{CTD} crystal structure, shown as a green mesh contoured at 1σ . **f**, Cartoon representation of one of the two CTD monomers present in the asymmetric unit.



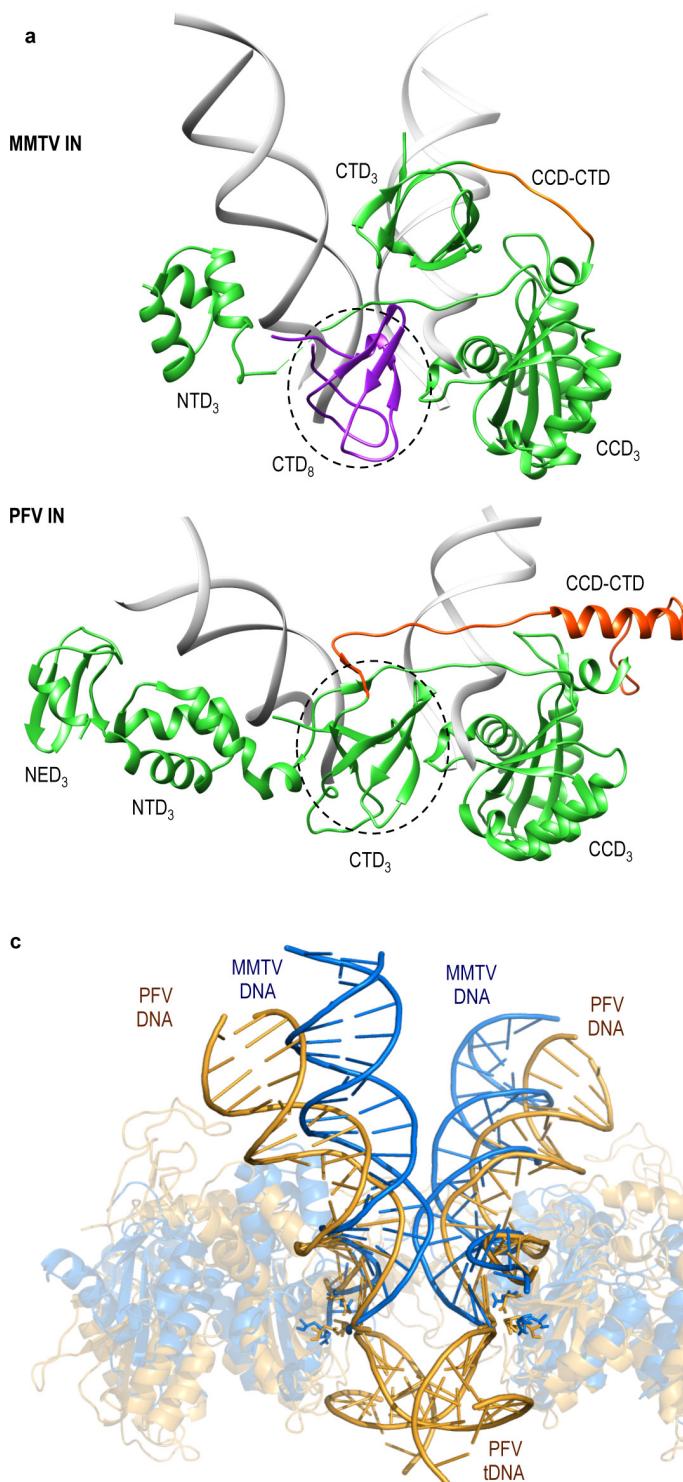
Extended Data Figure 6 | Molecular modelling of cryo-EM density.

a, Stereo views showing comparisons between the starting X-ray domain models and refined cryo-EM domain models for IN₁ highlight relatively minor structural perturbations that are evident only in the most flexible regions of the intasome. **b**, Linker region snapshots. Atomic models were built *de novo* from the cryo-EM density for the indicated linkers in the top two panels (residues 45–54 connecting NTD₁ and CCD₁ and CCD–CTD residues 211–213). Linkers NTD₂–CCD₂, CCD₅–CTD₅ and CCD₆–CTD₆ were not modelled, but are shown as cryo-EM density (red) in the lower panels. **c**, Stereo view of the cryo-EM model for the MMTV intasome core region (Extended Data Fig. 2d), generated using Rosetta^{15–17}. All domains

were refined starting with the X-ray crystal structures (Extended Data Fig. 5). Specific linker regions were built *de novo* (continuous red lines) from the cryo-EM density, whereas lower-resolution linker regions (red dotted lines) were omitted from the model. **d**, Fourier shell correlation curve between the refined cryo-EM core intasome model and map, showing an average resolution of 4.8 Å. **e**, Comparison of two NTD–CCD conformations in the intasome highlights the NTD–CCD linker, which assumes a retracted state in the outer IN₂ and IN₄ monomers of core intasome dimers A and B, respectively, as well as in flanking IN dimers C and D (left). The linker extends in core IN molecules IN₁ and IN₃, which interact with the vDNA (right).



Extended Data Figure 7 | Gel filtration profiles of IN_{WT} and IN mutant proteins. Elution profiles of mass standards in kilodaltons as well as theoretical protein monomer (M) and dimer (D) positions are indicated.



Extended Data Figure 8 | Comparisons of PFV and MMTV intasome structures. **a**, Cartoon representations of the inner IN₃ green subunits of the MMTV and PFV intasomes (Fig. 3a; vDNA strands are in grey). CCD-CTD linker regions are highlighted in orange, and dashed lines circle analogously positioned CTDs. Of note, this CTD in the MMTV structure is coloured differently because it originates from a separate IN molecule (IN₈ from flanking dimer D). **b**, Lengths of NTD-CCD and CCD-CTD interdomain linker regions across retroviral IN proteins; '+' indicates the presence of an N-terminal extension domain (NED). The multimeric state

of IN in known intasome structures is indicated by bold type. **c**, The PFV intasome with bound tDNA (PDB accession number 3OS2; orange) was superimposed with the MMTV intasome (blue). The distance between overlaid active sites is in each case ~26 Å. **d**, Ninety-degree rotation of superimposed structures, with proteins omitted for clarity. Canonical B-form tDNA (magenta) was superimposed with PFV intasome tDNA. The positions of phosphodiester bonds staggered by 4 bp in the PFV crystal structure or by 6 bp in the modelled tDNA are indicated by spheres.

Extended Data Table 1 | X-ray crystallography data collection and refinement statistics

Construct	CCD	NTD-CCD	CTD
Data collection			
Space group	P1	P12 ₁ 1	C222 ₁
Cell dimensions			
a, b, c (Å)	51.89, 53.71, 69.65	54.37, 83.15, 141.14	35.99, 42.28, 139.09
a, b, g (°)	69.69, 82.08, 63.97	90, 90.19, 90	90, 90, 90
Resolution (Å)*	46.6 - 1.70 (1.73 - 1.70)	70.6 - 2.72 (2.79 - 2.72)	40.4 - 1.50 (1.53 - 1.50)
R_{merge}	0.060 (0.57)	0.08 (0.534)	0.043 (0.585)
$\text{I}/\text{I}_{\text{l}}$	21.0 (2.0)	9.5 (2.0)	29.2 (3.8)
Completeness (%)	99.1 (95.6)	99.3 (99.0)	99.8 (99.9)
Redundancy	5.2 (2.8)	3.2 (3.1)	12.2 (8.9)
Refinement			
Resolution (Å)	32.8 - 1.70	70.6 - 2.72	40.4 - 1.50
No. reflections used	69,075	32,115	17,448
$R_{\text{work}}/R_{\text{free}}$	0.189/0.222	0.245/0.266	0.165/0.202
No. atoms			
Protein	4,983	9,110	890
Ligand/ion	0	12	8
Water	437	0	69
B-factors			
Protein	26.0	70.9	28.5
Ligand/ion	-	45.6	46.4
Water	33.5	-	46.9
R.m.s deviations			
Bond lengths (Å)	0.007	0.010	0.005
Bond angles (°)	0.954	1.281	0.911

*Data for the highest resolution shells are given in parenthesis.

Extended Data Table 2 | Cryo-EM data statistics

Construct	core MMTV intasome	full MMTV intasome
EM data collection/processing		
Microscope	Titan Krios	Titan Krios
Voltage	300	300
Camera	Gatan K2 Summit	Gatan K2 Summit
Defocus range (μm)	1.0-4.0	1.0-4.0
Defocus mean ± std (μm)	2.6 ± 0.6	2.6 ± 0.6
Exposure time (s)	10	10
Dose rate (e-/pixel/s)	6.9	6.9
Total dose (e-/Å ²)	40	40
Pixel size (Å)	1.31	1.31
Number of micrographs	2,714	2,714
Number of particles (processed)	147,850	147,850
Number of particles (refined)	77,365	77,365
Number of particles (in final map)	30,307	41,475
Symmetry	C2	C2
Resolution (global) (Å)*	4.8	6.0
Resolution range (local) (Å)	4 – 5	5 – 6
Map sharpening B-factor (Å ²)	-300	-460
Model refinement		
Space group	P1	-
Cell dimensions		
$a = b = c$ (Å)	151.2	-
$a = b = g$ (°)	90	-
Number of atoms (modeled)	11,462	-
Validation		
MolProbity score	1.46 (96 th percentile)	-
Clashscore, all atoms	2.27 (99 th percentile)	-
Protein		
Ramachandran favored (%)	1,115 (92.76)	-
allowed (%)	87 (7.24)	-
Disallowed (%)	0 (0)	-
Good rotamers (%)	1,035 (99.71)	-
Cβ deviations >0.25Å (%)	0 (0)	-
Cis Prolines (%)	8 / 88 (9.09)	-
Bad bonds (%)	2 / 10,140 (0.02)	-
Bad angles (%)	3 / 13,810 (0.02)	-
DNA		
Bad bonds (%)	0 / 1,834 (0)	-
Bad angles (%)	1 / 2,822 (0.04)	-
r.m.s. deviations		
Bond lengths (Å)	0.012	-
Bond angles (°)	1.334	-

*Resolution assessment based on frequency-limited refinement using the 0.143-threshold for resolution analysis.

Crystal structure of the Rous sarcoma virus intasome

Zhiqi Yin^{1,2,3*}, Ke Shi^{1,2,3*}, Surajit Banerjee⁴, Krishan K. Pandey⁵, Sibes Bera⁵, Duane P. Grandgenett⁵ & Hideki Aihara^{1,2,3}

Integration of the reverse-transcribed viral DNA into the host genome is an essential step in the life cycle of retroviruses. Retrovirus integrase catalyses insertions of both ends of the linear viral DNA into a host chromosome¹. Integrase from HIV-1 and closely related retroviruses share the three-domain organization, consisting of a catalytic core domain flanked by amino- and carboxy-terminal domains essential for the concerted integration reaction. Although structures of the tetrameric integrase–DNA complexes have been reported for integrase from prototype foamy virus featuring an additional DNA-binding domain and longer interdomain linkers^{2–5}, the architecture of a canonical three-domain integrase bound to DNA remained elusive. Here we report a crystal structure of the three-domain integrase from Rous sarcoma virus in complex with viral and target DNAs. The structure shows an octameric assembly of integrase, in which a pair of integrase dimers engage viral DNA ends for catalysis while another pair of non-catalytic integrase dimers bridge between the two viral DNA molecules and help capture target DNA. The individual domains of the eight integrase molecules play varying roles to hold the complex together, making an extensive network of protein–DNA and protein–protein contacts that show both conserved and distinct features compared with those observed for prototype foamy virus integrase. Our work highlights the diversity of retrovirus intasome assembly and provides insights into the mechanisms of integration by HIV-1 and related retroviruses.

Integrases (INs) from lentiviruses including HIV-1 and the phylogenetically closely related alpharetroviruses including avian Rous sarcoma virus (RSV) share the conserved three-domain organization consisting of the Zn²⁺-coordinating amino (N)-terminal domain (NTD), the catalytic core domain (CCD), and the β-strand-rich carboxy (C)-terminal domain (CTD) (Fig. 1a and Extended Data Fig. 1). We used the three-domain RSV IN construct biochemically fully active in concerted integration⁶ and a branched DNA substrate mimicking the product of the concerted integration reaction⁷ (Fig. 1b) to assemble and crystallize the stable RSV intasome complex (Extended Data Figs 2 and 3). The crystallized RSV intasome showed in solution an apparent molecular mass of 255 kDa by size-exclusion chromatography (SEC) and 240 (±10) kDa by SEC with multi-angle light scattering (SEC–MALS) analysis, larger than the expected mass of 168 kDa for a tetramer of RSV IN(1–270) bound to the branched DNA (Extended Data Fig. 4a–c). The structure of the RSV intasome was determined by molecular replacement phasing and refined to 3.8 Å resolution (Extended Data Table 1).

In contrast to the general assumption that retrovirus intasomes consist of an IN tetramer, the RSV intasome structure shows that it contains eight IN molecules (Fig. 1c–f and Supplementary Video), clearly supported by the selenium anomalous difference Fourier peaks for a crystal grown with selenomethionine-labelled IN (Extended Data Fig. 5). The observed octameric assembly (288 kDa) is consistent with the larger apparent molecular mass of RSV intasome *in vitro* and chemical cross-linking analysis (Extended Data Fig. 4d). The RSV IN octamer

contains a core tetramer, which consists of two sets of ‘proximal’ IN dimers, and two additional sets of ‘distal’ IN dimers. The CCDs of the four IN dimers are positioned in an approximately twofold symmetrical arrangement that resembles a parallelogram (Fig. 1c and d). The proximal IN dimer consists of ‘inner’ and ‘outer’ IN subunits, where the active site of the inner subunit accommodates the viral/target DNA junction (Fig. 2a–c). The NTD of each inner IN subunit interacts in *trans* with the viral DNA engaged by the opposing proximal IN dimer, with the extended linker between NTD and CCD traversing the grooves of both viral DNA molecules to make additional contacts (Figs 2a and 3a–c). The domain-swapped arrangement of NTD and CCD of the proximal IN dimers is analogous to that observed in the tetrameric prototype foamy virus (PFV) intasome².

However, despite the shared structural features that support the basic chemistry of integration, the RSV and PFV intasomes have very different architectures (Fig. 3 and Extended Data Fig. 6). The PFV intasome consists of four IN molecules, corresponding only to the core tetramer (two proximal IN dimers) of the RSV intasome. In the PFV intasome the inner IN subunits make all DNA interactions, while only the CCD is ordered for the outer subunits (Fig. 3d–f). In the RSV intasome, CTDs of the inner and outer subunits of the proximal IN dimer take unique conformations and tightly associate with each other (Fig. 2c), and this CTD dimer makes viral DNA contacts both in *cis* and in *trans* (Fig. 2a). Moreover, the octameric RSV intasome contains two additional IN dimers; the CTDs from the distal IN dimers bridge between the proximal IN dimers, making additional contacts with both viral DNAs analogously to the CTD of the catalytic IN subunit in the PFV intasome (Fig. 3b, e). The CCDs of the distal IN dimers, anchored to the core of the RSV intasome through these CTD interactions (Fig. 2a), are positioned at the outer corners of the parallelogram, loosely associated with the distal regions of target DNA through non-specific interactions (Figs 1d–f and 2a). The six remaining NTDs from the outer subunits of the proximal IN dimers and both subunits of the distal IN dimers are bound intra-molecularly to CCD (Extended Data Fig. 7c, d). In total, over 10,000 Å² of molecular surface is buried in IN–IN interfaces within the RSV intasome, approximately half of which is accounted for by the conserved CCD dimerization interface, and ~6,000 Å² in IN–DNA interfaces.

Although the CCDs and CTDs from all four IN dimers within the RSV intasome individually self-dimerize in the same fashions, the relative positioning of the CTDs with respect to the CCDs differs between the proximal and distal IN dimers, corresponding to their different roles. The CCD–CTD configuration for the proximal IN dimers is very similar to that previously observed for DNA-free RSV INs^{6,8} (Fig. 2c), suggesting that RSV IN dimer in its native conformation is poised for viral DNA binding and catalysis, as noted earlier⁶. The CTD of the inner catalytic subunit of the proximal dimer binds to viral DNA near the viral/target junction but on the opposite face of the double-stranded DNA, while it also makes contacts in *trans* with the other viral DNA (Fig. 2a, b). The CTD of the outer subunit of the proximal dimer binds to the distal region of the viral DNA in *cis*.

¹Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, Minneapolis, Minnesota 55455, USA. ²Institute for Molecular Virology, University of Minnesota, Minneapolis, Minnesota 55455, USA. ³Masonic Cancer Center, University of Minnesota, Minneapolis, Minnesota 55455, USA. ⁴Northeastern Collaborative Access Team, Cornell University, Advanced Photon Source, Lemont, Illinois 60439, USA. ⁵Institute for Molecular Virology, St. Louis University Health Sciences Center, St. Louis, Missouri 63104, USA.

*These authors contributed equally to this work.

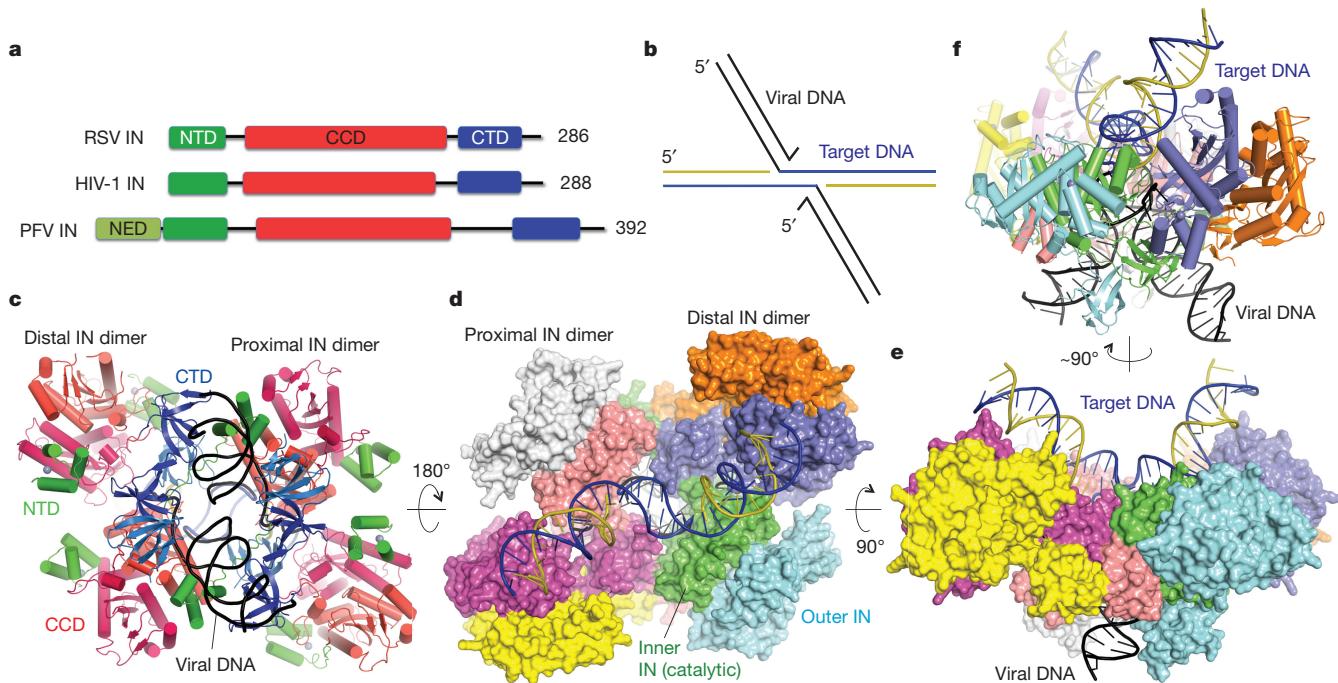


Figure 1 | Overall structure of the RSV intasome. **a**, Comparison of the domain organization between RSV, HIV-1, and PFV INs. **b**, Branched DNA structure mimicking the product of the concerted integration reaction used in assembling the RSV intasome. **c**, Structure of the RSV intasome, viewed along its pseudo-twofold axis from the viral DNA side. The three structural domains of IN are colour coded as in **a**. The grey

spheres represent zinc ions in NTD. Two subunits within each IN dimer are coloured slightly differently from each other. **d–f**, Structure of the RSV intasome with the eight IN molecules coloured individually, shown in three orthogonal orientations. The protein surface is shown in **d** and **e**. The DNA strands are colour coded as in **b**.

Unlike in the proximal IN dimer, the CCD–CTD configuration for the distal IN dimers shows deviation from the canonical conformation,

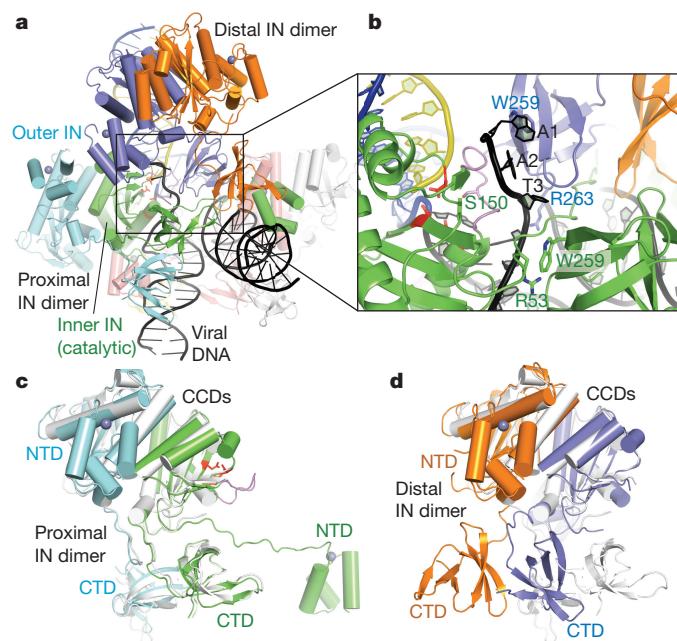


Figure 2 | Proximal and distal IN dimers. **a**, View highlighting interactions between a proximal (green/cyan) and a distal (slate/orange) IN dimers and their interactions with DNA. The IN and DNA strands are coloured as in Fig. 1d–f. **b**, Close-up view around the viral DNA terminus. The CCD loop (residues 144–153) centred around Ser150 of the catalytic IN subunit is coloured in violet. The catalytic triad residues (D64, D121, E157) are shown in red sticks in **a–c**. **c, d**, Superposition of the proximal IN dimer (**c**; green/cyan) or the distal IN dimer (**d**; slate/orange) on the RSV IN CCD–CTD dimer in its native conformation (PDB accession number 4WF1; light grey)⁶. The grey spheres represent zinc ions in NTD.

which can be described by a swing of the CTDs relative to the CCDs and disruption of the parallel β -sheet-like conformation of the CCD–CTD linkers⁸ (Fig. 2d). This alternative CCD–CTD orientation allows the distal IN dimers to fit in the intasome without making steric clashes with the proximal INs or the 5' overhang of the viral DNA strand (Fig. 2a). The CCDs of the distal IN dimers may be positioned similarly in the absence of target DNA, serving as a platform for target DNA capturing.

The asymmetrically associated CTDs of the proximal and distal IN dimers further interact with each other and with the NTD of the catalytic IN subunit to crosslink between the two viral DNAs (Figs 2a and 4a, b). Each CTD dimer interacts with both of the viral DNA molecules at different positions, resulting in four distinct DNA-binding modes of individual CTDs. The basic amino acids Arg227, Arg244, Arg263, and Lys266 from various CTD monomers make contacts with the viral DNA molecules, and each viral DNA is sandwiched between separate CTD dimers (Fig. 4c). Arg244 of the inner catalytic subunit of the proximal IN dimer is positioned in the major groove of viral DNA, closest to G7 of the non-transferred strand. The GC pair at this position is critical for concerted integration by RSV IN⁹. The corresponding residue Glu246 of HIV-1 IN was shown by disulfide cross-linking studies to interact with A7 of the non-transferred strand¹⁰, suggesting similar modes of viral DNA sequence recognition by CTD between RSV and HIV-1 INs. Both RSV R244A/C and HIV-1 E246A IN mutants show reduced 3' processing and strand-transfer activities^{6,11,12}, possibly reflecting the importance of these residues from various IN subunits. Mutation of conserved Trp233, which is stacked between the Arg227 and Lys266 side chains, to Glu or Ala but not Phe, abolishes binding to the viral DNA long terminal repeat sequence and concerted integration by RSV IN¹³. The corresponding HIV-1 IN mutations W235E/A/F have parallel effects on concerted integration activity and virus replication, suggesting the importance of an aromatic residue at this position in orienting the basic side chains^{14,15}. Similarly, mutation of Trp259 buried in the CTD dimer interface (Fig. 4b)^{6,8} as well as involved in multiple interactions near the viral DNA 5' end (Fig. 2b) abolishes all enzymatic

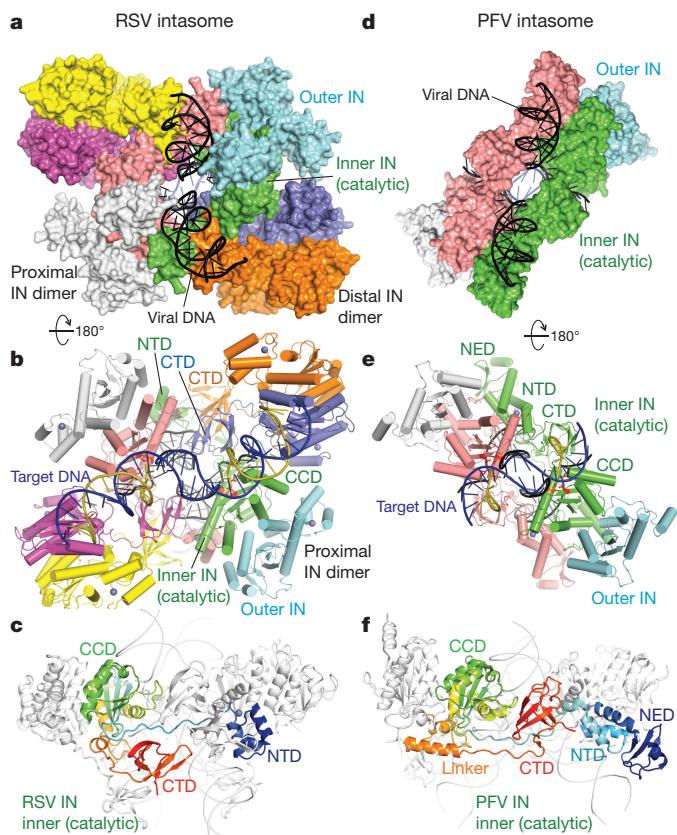


Figure 3 | Comparison between the RSV and PFV intasomes. **a, b**, The octameric RSV intasome. The eight IN molecules and the DNA strands are coloured as in Fig. 1d–f. The catalytic triad (DDE) residues of the inner subunit of the proximal IN dimers are shown in red in **b**. **c**, Conformation of the inner catalytic IN subunit in the RSV intasome. The protein chain is coloured in a gradient of blue to red from N to C termini, respectively. The NTDs and CCDs of the distal IN dimer are omitted. **d, e**, The tetrameric PFV intasome³. The colour scheme follows that used for the proximal IN dimers of RSV intasome in **a** and **b**. The direction of DNA helical axis between the two integration sites on opposing strands (6 bp for RSV and 4 bp for PFV) differs significantly between the RSV (**b**) and PFV (**e**) intasomes. **f**, The inner catalytic IN subunit in the PFV intasome, coloured in a gradient of blue to red as in **c**. Note the presence of an extra domain on the N terminus (NED) and different positioning of CTD compared with RSV IN in **c**.

activities of RSV IN^{6,16}, reflecting the important roles of this residue in engaging viral DNA.

The CCD of the catalytic RSV IN molecules engages viral and target DNAs primarily through interactions in their minor grooves, as observed for PFV IN^{2,3}. The long α -helix (α 7; residues 154–174) that harbours one of the catalytic triad of metal-coordinating residues, Glu157, inserts into a significantly widened minor groove near the viral DNA terminus, where Arg158, Arg161, and Lys164 side chains make DNA base or backbone contacts (Fig. 4d and Extended Data Fig. 7h, i). The preceding loop centred around Ser150, which is highly flexible in the DNA-free IN^{6,8,17}, is positioned between the viral DNA 5' overhang and the 3' end of the cleaved target DNA, displacing the T3 base opposite the terminal adenine (A20) of the transferred strand (Fig. 2b). DNA contacts in this minor groove also include hydrogen bonding between Thr66 and the backbone phosphate group of the terminal nucleotide (A20) of the transferred viral DNA strand (Fig. 4d). Mutations of the corresponding HIV-1 IN residue T66A/I/K confer resistance against the IN strand-transfer inhibitors (INSTIs; <http://hivdb.stanford.edu/DR/INIResNote.html>), probably as a result of subtle changes of the IN–viral DNA interface. The NTD of the opposing catalytic subunit, contributed in *trans*, binds in the adjacent major

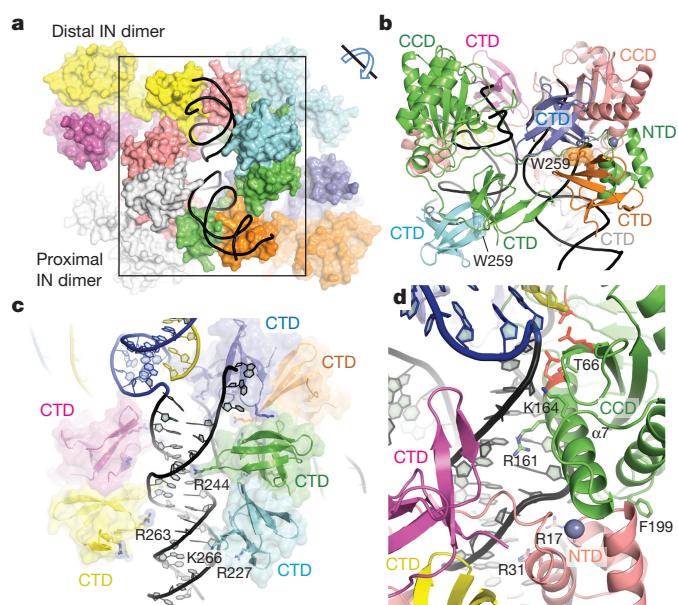


Figure 4 | Viral DNA contacts. **a**, Surface representation of the RSV intasome similar to Fig. 3a. **b**, A crown-like structure formed by the CTD of all eight INs and NTD and CCD of the catalytic INs, which encircles the two viral DNA molecules in black (boxed region in **a**). Trp259 side chains buried in the interface of each CTD dimer are shown. **c**, Viral DNA contacts by the CTDs. The side chains of Arg227, Arg244, Arg263, and Lys266 are shown. NTD and CCD are omitted. **d**, Close-up view around the viral DNA end, showing interactions by CCD of a catalytic subunit, NTD of the opposing catalytic subunit contributed in *trans*, and the CTDs of a distal IN dimer. The catalytic triad residues are shown in red sticks. The grey sphere represents a zinc ion in NTD.

groove and places Arg17 and Arg31 side chains for potential base-specific contacts (Fig. 4d). The hydrophobic NTD–CCD interface is centred on Phe199 from CCD, which explains why the F199K mutation selectively abolishes concerted integration by RSV IN^{6,16}. The CTDs of the distal IN dimer further extend the viral DNA interactions in this region (Fig. 4d).

The three flipped-out terminal nucleotides at the 5' end of the non-transferred viral DNA strand, including the two overhanged bases and the displaced T3, bridge between CCD of the catalytic IN subunit and a CTD from the distal IN dimer, rather than between CCD and CTD of the same molecule as observed in the PFV intasome (Figs 2b, 4b and 5a). Arg263 from this CTD points towards the non-transferred strand G4 opposite the CA dinucleotide at the viral DNA terminus (Fig. 2b), which could be relevant to the effect of mutation of the corresponding HIV-1 residue Arg263 on catalysis and drug resistance¹⁸. The two viral DNAs in the RSV intasome branch out with their helical axes skewed and at an angle of $\sim 60^\circ$, which is smaller than the viral DNA split of $\sim 80^\circ$ in the PFV intasome (Extended Data Fig. 8). Accordingly, the viral DNA molecules in the RSV intasome are positioned closer to each other than those in the PFV intasome, with the backbone phosphate oxygen atoms at the closest point ~ 5 Å apart. The viral DNA molecules in the RSV intasome are surrounded by a highly positively charged surface formed by a network of CTDs and NTDs, which may alleviate potential electrostatic repulsions between DNA strands and help hold the complex together (Fig. 4a–c).

The target DNA in the RSV intasome shows a strong overall bending of $\sim 90^\circ$ away from the core of the complex (Fig. 5a and Extended Data Fig. 7b), which may help prevent the reversal of integration as noted for the PFV intasome and related transpososome structures^{3,19}. The bent conformation is stabilized by the minor groove contacts made by CCD of the catalytic subunits near the viral/target DNA junction, which include insertion of a short helix (α 5) harbouring Ser124, a residue important in target DNA capturing²⁰. Localized kinks at the

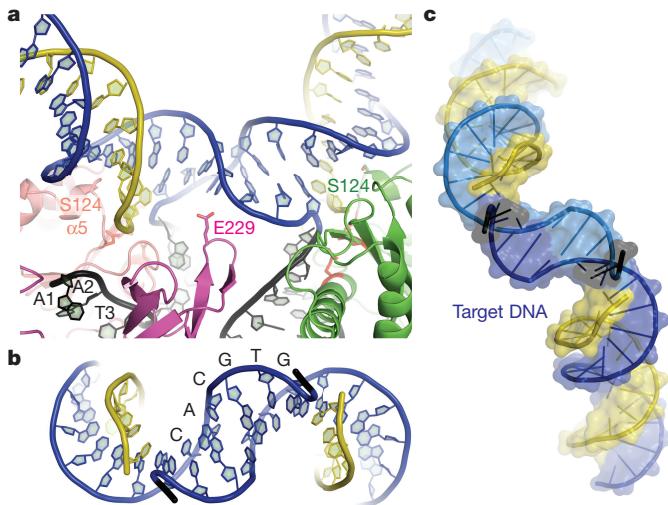


Figure 5 | Target DNA. **a**, The sharply bent target DNA spanning the CCDs of the catalytic IN subunits. Ser124 and Glu229 side chains potentially involved in target DNA contacts are shown. **b**, The central 6 bp region viewed from the viral DNA side, showing distorted DNA conformation with severely compromised base-stacking. **c**, The target DNA viewed from the viral DNA side, highlighting a large shift in the helical axis.

viral/target DNA junctions cause the DNA trajectory also to zigzag in the plane perpendicular to the direction of the primary bending, creating a ~ 20 Å shift in the helical axis with an overall positive writhe (Fig. 5c). As a result, the target DNA conformation in the RSV intasome deviates significantly from that in the PFV intasome³ (Extended Data Fig. 8c) or the DNA structure in nucleosomes²¹. This suggests that RSV integration into nucleosomes would require a large conformational change in the target DNA, probably more extensive than that observed in the PFV intasome–nucleosome complex⁴.

The active sites of the catalytic IN subunits in the RSV intasome are separated by ~ 27 Å, larger than the distance of ~ 21 Å across the major groove between the backbone phosphates 6 base pairs (bp) apart in a B-form DNA. Accordingly, the 6 bp spacer region of target DNA is underwound and shows severely compromised base-stacking in addition to widening of the major groove (Fig. 5b and Extended Data Fig. 9). In particular, the central dinucleotide step shows a distorted conformation where the unstacked nucleobases stack on the deoxyribose moiety of the adjacent nucleotide; the unique conformation may explain the modest target sequence preference in RSV integration^{22,23}. A CTD from each distal IN dimer has a loop (between $\beta 6$ and $\beta 7$) positioned over the major groove of the target DNA in this region, with the Glu229 side chain positioned for potential base contacts (Fig. 5a). This loop from various CTDs is involved in viral DNA contacts, interactions between NTD and CTDs, and those linking the CTDs from the proximal and distal IN dimers in the RSV intasome (Fig. 4b and Extended Data Fig. 7e–g).

The RSV intasome structure shows a novel architecture of IN–DNA complex and highlights a remarkable diversity among retrovirus intasome assemblies. In particular, the structurally conserved CTD plays critical roles in the octameric RSV intasome distinct from those played by CTD in the tetrameric PFV intasome^{2,3,5}. PFV IN features a unique ~ 30 -amino-acid insertion between CCD and CTD (Fig. 1a and Extended Data Fig. 1), and the resulting long interdomain linker not only makes direct viral DNA contacts but also allows CTD of the catalytic IN subunit to fit between the NTD and CCD from the same molecule to interact with both viral DNAs (Fig. 3f). In case of RSV IN with a much shorter CCD–CTD linker, the NTD and CCD of the catalytic IN molecule is instead bridged by the CTDs contributed in *trans* from the distal IN dimer (Fig. 3c). Analogously, the extra ~ 50 amino acids on the N terminus

of PFV IN constitute an independent DNA-binding domain (NED), and its interaction with the distal region of the viral DNA substitutes for viral DNA interactions made by the multiple copies of CTDs as observed in the RSV intasome (Fig. 3a, d). The larger (408 amino acids) IN from a gammaretrovirus murine leukaemia virus also has both the NED and a long CCD–CTD linker²⁴, consistent with the idea that these two structural features have complementary functions. The different modes of viral DNA engagement by either tetrameric or octameric IN, using similar sets of structurally conserved domains, suggest divergent evolution of the integration machineries. Of note, an octameric intasome assembly similar to that reported here has recently been observed by cryo-electron microscopy for another three-domain IN from mouse mammary tumour virus²⁵.

The integration of retroviral DNA into the host genome was first postulated on the basis of the observation of RSV-infected cells²⁶, and identification and characterization of IN from retroviruses including RSV have provided a basic understanding of retrovirus integration and foundation for studying HIV-1 IN^{27–30}. Our findings on the RSV intasome structure shed light on the previously unappreciated diversity in retroviral DNA integration and define a new framework for future studies of retrovirus IN. Availability of the intasome structures from multiple retrovirus systems will also allow more accurate modeling of the HIV-1 IN–DNA interactions to help development of novel antiretroviral drugs, including IN inhibitors that target outside the most conserved active-site features.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 August; accepted 23 December 2015.

1. Craigie, R. & Bushman, F. D. HIV DNA integration. *Cold Spring Harb. Perspect. Med.* **2**, a006890 (2012).
2. Hare, S., Gupta, S. S., Valkov, E., Engelman, A. & Cherepanov, P. Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature* **464**, 232–236 (2010).
3. Maertens, G. N., Hare, S. & Cherepanov, P. The mechanism of retroviral integration from X-ray structures of its key intermediates. *Nature* **468**, 326–329 (2010).
4. Maskell, D. P. *et al.* Structural basis for retroviral integration into nucleosomes. *Nature* **523**, 366–369 (2015).
5. Gupta, K. *et al.* Solution conformations of prototype foamy virus integrase and its stable synaptic complex with U5 viral DNA. *Structure* **20**, 1918–1928 (2012).
6. Shi, K. *et al.* A possible role for the asymmetric C-terminal domain dimer of Rous sarcoma virus integrase in viral DNA binding. *PLoS ONE* **8**, e56892 (2013).
7. Yin, Z., Lapkouski, M., Yang, W. & Craigie, R. Assembly of prototype foamy virus strand transfer complexes on product DNA bypassing catalysis of integration. *Protein Sci.* **21**, 1849–1857 (2012).
8. Yang, Z. N., Mueser, T. C., Bushman, F. D. & Hyde, C. C. Crystal structure of an active two-domain derivative of Rous sarcoma virus integrase. *J. Mol. Biol.* **296**, 535–548 (2000).
9. Vora, A., Bera, S. & Grandgenett, D. Structural organization of avian retrovirus integrase in assembled intasomes mediating full-site integration. *J. Biol. Chem.* **279**, 18670–18678 (2004).
10. Gao, K., Butler, S. L. & Bushman, F. Human immunodeficiency virus type 1 integrase: arrangement of protein domains in active cDNA complexes. *EMBO J.* **20**, 3565–3576 (2001).
11. Peletskaya, E. *et al.* Localization of ASV integrase–DNA contacts by site-directed crosslinking and their structural analysis. *PLoS ONE* **6**, e27751 (2011).
12. Lutzke, R. A. & Plasterk, R. H. Structure-based mutational analysis of the C-terminal DNA-binding domain of human immunodeficiency virus type 1 integrase: critical residues for protein oligomerization and DNA binding. *J. Virol.* **72**, 4841–4848 (1998).
13. Chiu, R. & Grandgenett, D. P. Molecular and genetic determinants of rous sarcoma virus integrase for concerted DNA integration. *J. Virol.* **77**, 6482–6492 (2003).
14. Chen, H., Wei, S. Q. & Engelman, A. Multiple integrase functions are required to form the native structure of the human immunodeficiency virus type I intasome. *J. Biol. Chem.* **274**, 17358–17364 (1999).
15. Li, M. & Craigie, R. Processing of viral DNA ends channels the HIV-1 integration reaction to concerted integration. *J. Biol. Chem.* **280**, 29334–29339 (2005).
16. Bojja, R. S. *et al.* Architecture of a full-length retroviral integrase monomer and dimer, revealed by small angle X-ray scattering and chemical cross-linking. *J. Biol. Chem.* **286**, 17047–17059 (2011).

17. Lubkowski, J. *et al.* Atomic resolution structures of the core domain of avian sarcoma virus integrase and its D64N mutant. *Biochemistry* **38**, 13512–13522 (1999).
18. Quashie, P. K. *et al.* Characterization of the R263K mutation in HIV-1 integrase that confers low-level resistance to the second-generation integrase strand transfer inhibitor dolutegravir. *J. Virol.* **86**, 2696–2705 (2012).
19. Montaño, S. P., Pigli, Y. Z. & Rice, P. A. The Mu transpososome structure sheds light on DDE recombinase evolution. *Nature* **491**, 413–417 (2012).
20. Harper, A. L., Sudol, M. & Katzman, M. An amino acid in the central catalytic domain of three retroviral integrases that affects target site selection in nonviral DNA. *J. Virol.* **77**, 3838–3845 (2003).
21. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
22. Serrao, E., Ballandras-Colas, A., Cherepanov, P., Maertens, G. N. & Engelman, A. N. Key determinants of target DNA recognition by retroviral intasomes. *Retrovirology* **12**, 39 (2015).
23. Wu, X., Li, Y., Crise, B., Burgess, S. M. & Munroe, D. J. Weak palindromic consensus sequences are a common feature found at the integration target sites of many retroviruses. *J. Virol.* **79**, 5211–5214 (2005).
24. Aiyer, S. *et al.* Structural and sequencing analysis of local target DNA recognition by MLV integrase. *Nucleic Acids Res.* **43**, 5647–5663 (2015).
25. Ballandras-Colas, A. *et al.* Cryo-EM reveals a novel octameric integrase structure for β-retrovirus intasome function. *Nature* <http://dx.doi.org/10.1038/nature16955> (this issue).
26. Temin, H. M. The participation of DNA in Rous sarcoma virus production. *Virology* **23**, 486–494 (1964).
27. Grandgenett, D. P., Vora, A. C. & Schiff, R. D. A 32,000-dalton nucleic acid-binding protein from avian retrovirus cores possesses DNA endonuclease activity. *Virology* **89**, 119–132 (1978).
28. Donehower, L. A. & Varmus, H. E. A mutant murine leukemia virus with a single missense codon in pol is defective in a function affecting integration. *Proc. Natl. Acad. Sci. USA* **81**, 6461–6465 (1984).
29. Panganiban, A. T. & Temin, H. M. The retrovirus *pol* gene encodes a product required for DNA integration: identification of a retrovirus *int* locus. *Proc. Natl. Acad. Sci. USA* **81**, 7885–7889 (1984).
30. Schwartzberg, P., Colicelli, J. & Goff, S. P. Construction and analysis of deletion mutations in the *pol* gene of Moloney murine leukemia virus: a new viral function required for productive infection. *Cell* **37**, 1043–1052 (1984).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank K. Kurahashi for generating many mutant IN expression plasmids, J. Kankanala and Z. Wang for synthesizing Pt-modified oligonucleotides, and J. Nix for help with X-ray data collection. X-ray data were collected at the Advanced Photon Source (APS) NE-CAT beamlines, which are supported by the National Institute of General Medical Sciences (P41 GM103403). APS is a US Department of Energy Office of Science User Facility operated by Argonne National Laboratory under contract DE-AC02-06CH11357. This research was supported by National Institutes of Health grants GM109770, AI087098 to H.A. and AI100682 to D.P.G.

Author Contributions Z.Y. designed strategy and developed protocols for producing and crystallizing the RSV intasome, and prepared all crystals. Z.Y., K.S. and S.Ba. collected X-ray diffraction data. K.S. analysed the data and determined the crystal structure. K.P. and S.Ba. examined behaviours of the RSV intasome in solution and analysed mutant IN activities. D.P.G. and H.A. conceived the project and supervised the research. H.A. wrote the manuscript with important input from all authors.

Author Information Atomic coordinates and structure factors have been deposited in the Protein Data Bank (PDB) under accession number 5EJK. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.A. (aihar001@umn.edu).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

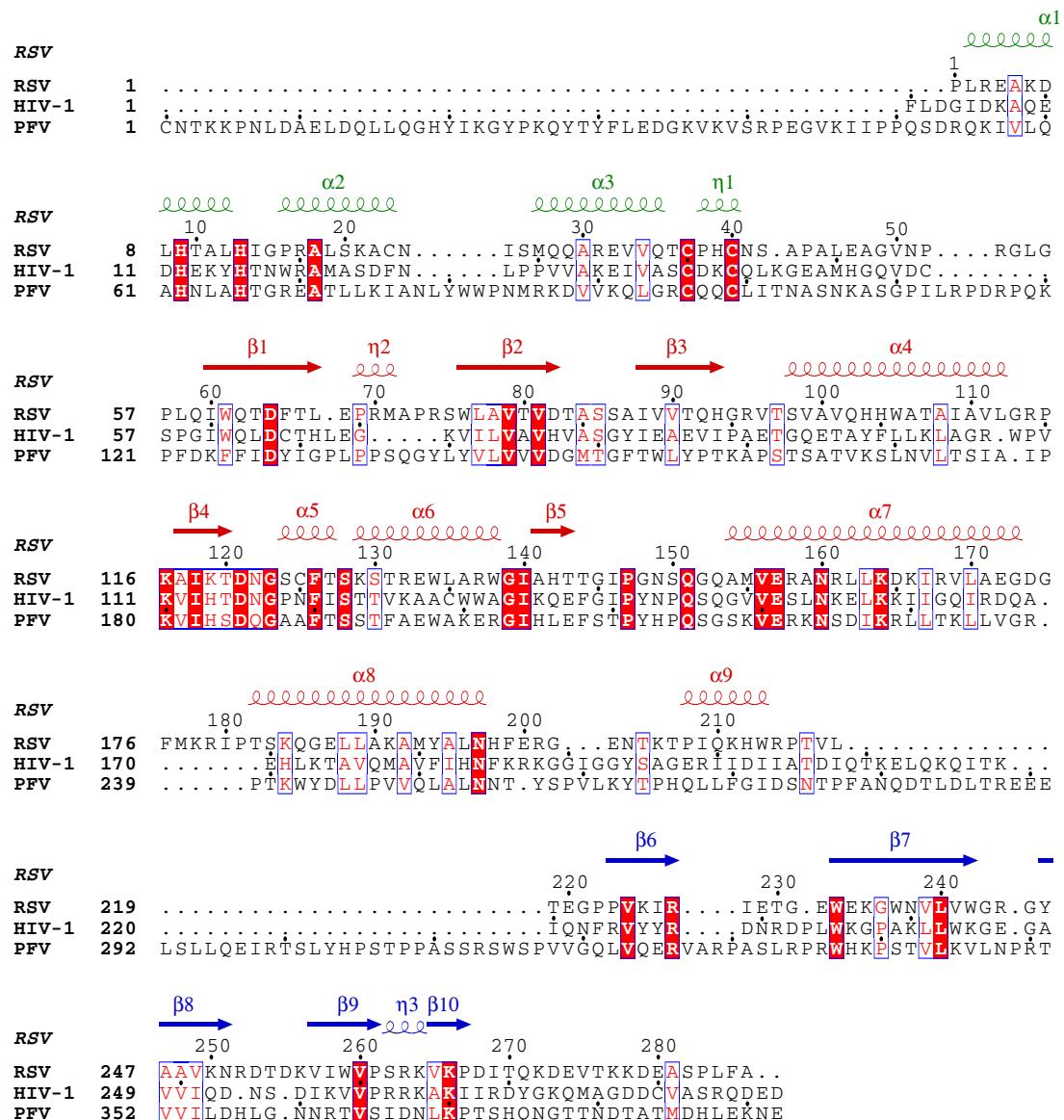
RSV intasome preparation. RSV IN and its mutant forms were overexpressed in *Escherichia coli* BL21 (DE3) and purified as previously described⁶. The proteins were stored in aliquots at -80°C in 20 mM HEPES-NaOH (pH 7.5), 1.0 M NaCl, 20 μM ZnCl₂, 5 mM 2-mercaptoethanol, and 10% (w/v) glycerol. The branched DNA substrate mimicking the product of the concerted integration reaction was obtained by annealing three synthetic oligonucleotides (Integrated DNA Technologies). A similar strategy had been used previously to prepare a PFV IN-DNA complex, which demonstrated that the IN-DNA complex assembled on the designed integration product has essentially an identical structure to the equivalent complex formed via the forward integration reaction⁷. The viral DNA branches carrying the high-affinity gain-of-function mutant RSV U3 long terminal repeat sequence (GU3)³¹ were attached to the target DNA duplex with a palindromic 6 bp spacer to generate a fully symmetrized structure. To prepare the RSV IN-DNA complexes for crystallization, 30 μM half-site DNA substrate was mixed with 120 μM RSV IN in 20 mM HEPES-NaOH (pH 7.5), 500 mM NaCl, 20% (w/v) glycerol, and 1 mM tris-(2-carboxyethyl)phosphine (TCEP). The mixture was dialysed against low salt buffer (20 mM HEPES-NaOH pH 7.5, 125 mM NaCl, 20% (w/v) glycerol, 1 mM TCEP) at 25°C overnight. At the end of this first dialysis, essentially all IN and IN-DNA complex precipitated. The mixture was subsequently dialysed against high salt buffer (20 mM MES-NaOH pH 6.0, 1.2 M NaCl, 20% (v/v) dimethyl sulfoxide (DMSO), 5% (w/v) glycerol, 1 mM TCEP) at 25°C for 2–4 h. At the end of this second dialysis, the reaction mixture became clear. The solubilized RSV IN-DNA complex (intasome) was purified through SEC (Superdex 200 10/300 GL, GE Healthcare) running with 20 mM MES-NaOH pH 6.0, 1.2 M NaCl, 20% (v/v) DMSO, 5% (w/v) glycerol, and 1 mM TCEP. The isolated intasome remains stable in the high-salt condition containing 1.2 M NaCl that precludes complex formation, suggesting that once the intasome forms, it is kinetically trapped. The solubility-enhancing RSV IN mutation F199K completely abolished intasome formation. For the SEC-MALS analysis, a modified condition was used for the intasome re-solubilization and isolation for better baseline stability (Extended Data Fig. 4c).

Crystallization. For crystallization, the RSV intasome assembled through dialysis and purified by SEC was concentrated to 4–6 mg ml⁻¹ using a centrifugal concentrator (Amicon). Various combinations of the lengths of the viral DNA (ranging from 16 bp to 25 bp) and flanking target DNA (ranging from 14 bp to 22 bp, corresponding to the full-target DNA lengths of 34–50 bp) were screened in crystallization trials. The extensive screening yielded only one crystal form with a specific combination of 22-base (length of the non-transferred strand) viral DNA branches and 16 bp target DNA flanks on either side of the central 6 bp spacer (Fig. 1b and Extended Data Fig. 2). DNA substrates with two slightly different target sequences were used (5'-AATGTTGCTTATGCAATACTC-3'/5'-GAGTATTGCTAAGAAGAACAGTGACAGAACATCTTGAAGACACT-3'/5'-AGTGTCTTCAAGATTTC-3', or 5'-AATGTTGCTTATGCAATACTC-3'/5'-GAGTATTGCTAAGACAAACA GTCGACCAACCTTCAACTTAGC-3'/5'-GCTAAGTTGAAGGTTG-3'), and they produced essentially the same crystals. The RSV intasome crystals were grown through reverse vapour diffusion in hanging drops at 22°C , by mixing 1.5 μl IN-DNA complex solution with 1.5 μl reservoir solution (3.2 M sodium formate). Crystals appeared within 3–5 days and reached a size of \sim 150–300 μm in 3–5 days. Even though the RSV intasome crystals initially diffracted X-ray poorly (\sim 10 \AA), soaking the crystals with a metatungstate cluster compound dramatically improved the resolution (Extended Data Fig. 3b, c). The tungsten cluster was later found to bind between IN dimers from separate intasome complexes to mitigate crystal lattice disorder. The crystals were soaked overnight with 0.15 mM metatungstate cluster Na₆[H₂W₁₂O₄₀] in a stabilization buffer consisting of 3.2 M sodium formate, 16 mM MES-NaOH pH 6.0, 0.8 M NaCl, 16% (v/v) DMSO, 4% (w/v) glycerol, and 1 mM TCEP. After soaking, the crystals were cryo-protected in 3.2 M sodium formate, 16 mM MES pH 6.0, 0.8 M NaCl, 16% (v/v) DMSO, 12% (w/v) glycerol, and 1 mM TCEP, and were frozen by rapid immersion in liquid nitrogen. The full-length wild-type RSV IN(1–286), the C-terminally truncated RSV IN(1–270), and its various mutant forms tested produced essentially the same crystals with indistinguishable X-ray diffraction properties.

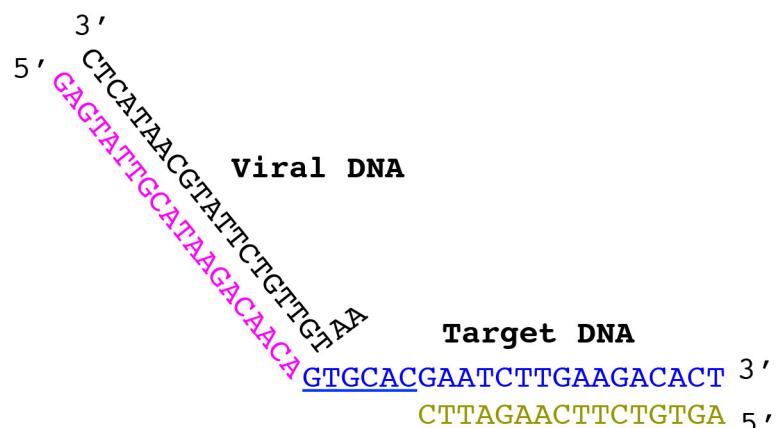
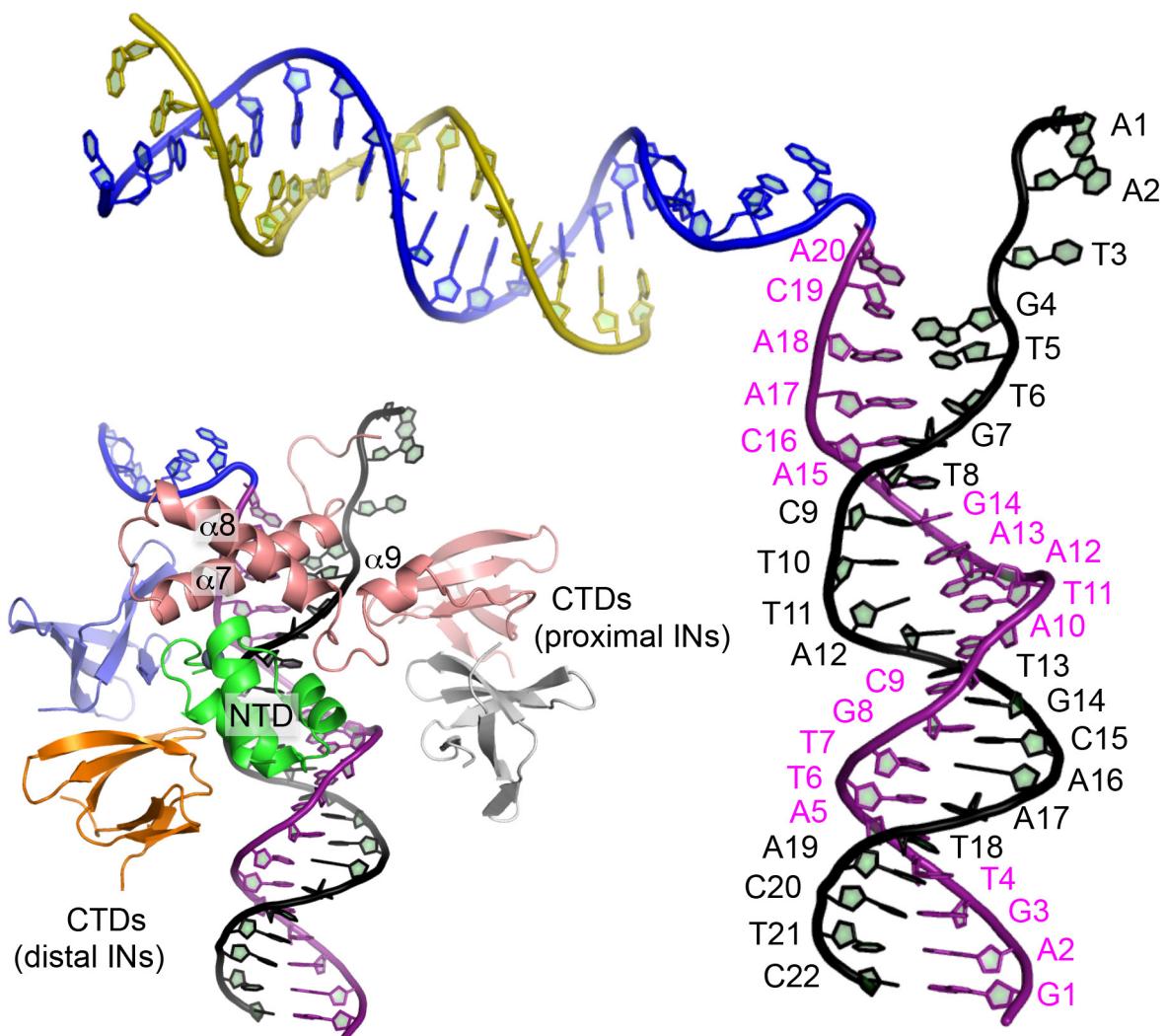
Structure determination. X-ray diffraction data were collected at the Advanced Photon Source Northeastern Collaborative Access Team beamlines (24-ID-C/E) and the Advanced Light Source Molecular Biology Consortium (4.2.2) beamline and processed using HKL2000 (ref. 32) or XDS³³. The RSV intasome crystals showed varying degrees of pseudo-merohedral twinning with twin operator

[l, -k, h], owing to the very similar a and c unit cell dimensions of the primitive monoclinic lattice. Thus, we screened a large number of crystals to identify ones that diffracted to higher resolution and had smaller twin fractions. The structure of RSV intasome was determined by molecular replacement with PHASER³⁴, using the RSV IN CCD, CTD (PDB accession number 4FW1)⁶, and a 16 bp B-form DNA as search models. Eight copies of CCD, one copy of CTD, and three copies of DNA molecules were located. Refinement of the partial model revealed electron density for two copies of the metatungstate clusters. The metatungstate clusters were placed into the electron density by molecular replacement using MOLREP³⁵. Subsequent iterative model building using COOT³⁶ and refinement with PHENIX suite³⁷ allowed placement of the remaining seven copies of CTD, eight copies of NTD generated using MODELLER³⁸ on the basis of HIV-1 NTD (PDB accession number 1K6Y)³⁹, and building of the inter-domain linkers as well as the remaining parts of the DNA molecule guided by the difference electron density maps. A third metatungstate cluster, more weakly bound than the first two, was positioned manually into residual density. The DNA base pairs and base-stacking restraints were used throughout the refinement. The geometry restraints for protein included the reference-model restraints for CCD and CTD based on the higher-resolution RSV IN structure (4FW1)⁶, and the secondary structure and zinc-coordination restraints for NTD. Atomic displacement parameters were refined with grouped B-factors per residue for protein and DNA, and a total of 54 TLS groups assigned by PHENIX³⁷. Twelve tungsten atoms representing each cluster were refined as a rigid body. The asymmetric unit of the crystal contains one complete RSV intasome, which includes eight IN molecules and two viral DNA branches emanating from a strongly bent 38 bp target DNA. The data set used for the final refinement was from an RSV intasome crystal grown using selenomethionine-labelled RSV IN(1–270) with the following amino-acid substitutions: C23S, L112M, L135M, L162M, L163M, L188M, and L189M, which we confirmed to be active in concerted integration and inhibited by INSTI similarly to the wild-type RSV IN, and a DNA substrate carrying a nick at the middle of each target DNA branch (the 16-base DNA strand shown in olive in Fig. 1b has a nick 8 bases from either end). The nick occasionally facilitated crystal growth, but it was not necessary for crystallization and did not change the space group or the unit cell parameters compared with crystals grown without the nick. Because this nick in the target DNA is not biologically relevant, it is not shown in Figs 1b, d–f, 3b and 5a, c to avoid confusion. Twin refinement protocol was not used as the data set used for the final refinement had a low (<10%) twin fraction. The summary of data collection and refinement statistics is shown in Extended Data Table 1. The paired-refinement procedure⁴⁰ was performed in steps of 0.1 \AA to determine the high-resolution limit (Extended Data Fig. 3f). The register of amino acids in the final model was verified by the selenium anomalous difference Fourier peaks (Extended Data Fig. 5). Ramachandran analysis shows that 96.0, 3.9, and 0.1% of the protein residues are in the most favoured, allowed, and disallowed regions, respectively. The NTD–CCD linker in some of the non-catalytic IN molecules and the last four base pairs at the distal end of one of the viral DNA molecules were not built owing to poor electron density. The molecular graphics images were produced using PYMOL (www.pymol.org).

31. Vora, A. C. *et al.* Avian retrovirus U3 and U5 DNA inverted repeats. Role of nonsymmetrical nucleotides in promoting full-site integration by purified virion and bacterial recombinant integrases. *J. Biol. Chem.* **272**, 23938–23945 (1997).
32. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
33. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
34. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
35. Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022–1025 (1997).
36. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
37. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
38. Šali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993).
39. Wang, J. Y., Ling, H., Yang, W. & Craigie, R. Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein. *EMBO J.* **20**, 7333–7343 (2001).
40. Karpilus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030–1033 (2012).
41. Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, W320–W324 (2014).
42. Pandey, K. K. *et al.* Rous sarcoma virus synaptic complex capable of concerted integration is kinetically trapped by human immunodeficiency virus integrase strand transfer inhibitors. *J. Biol. Chem.* **289**, 19648–19658 (2014).

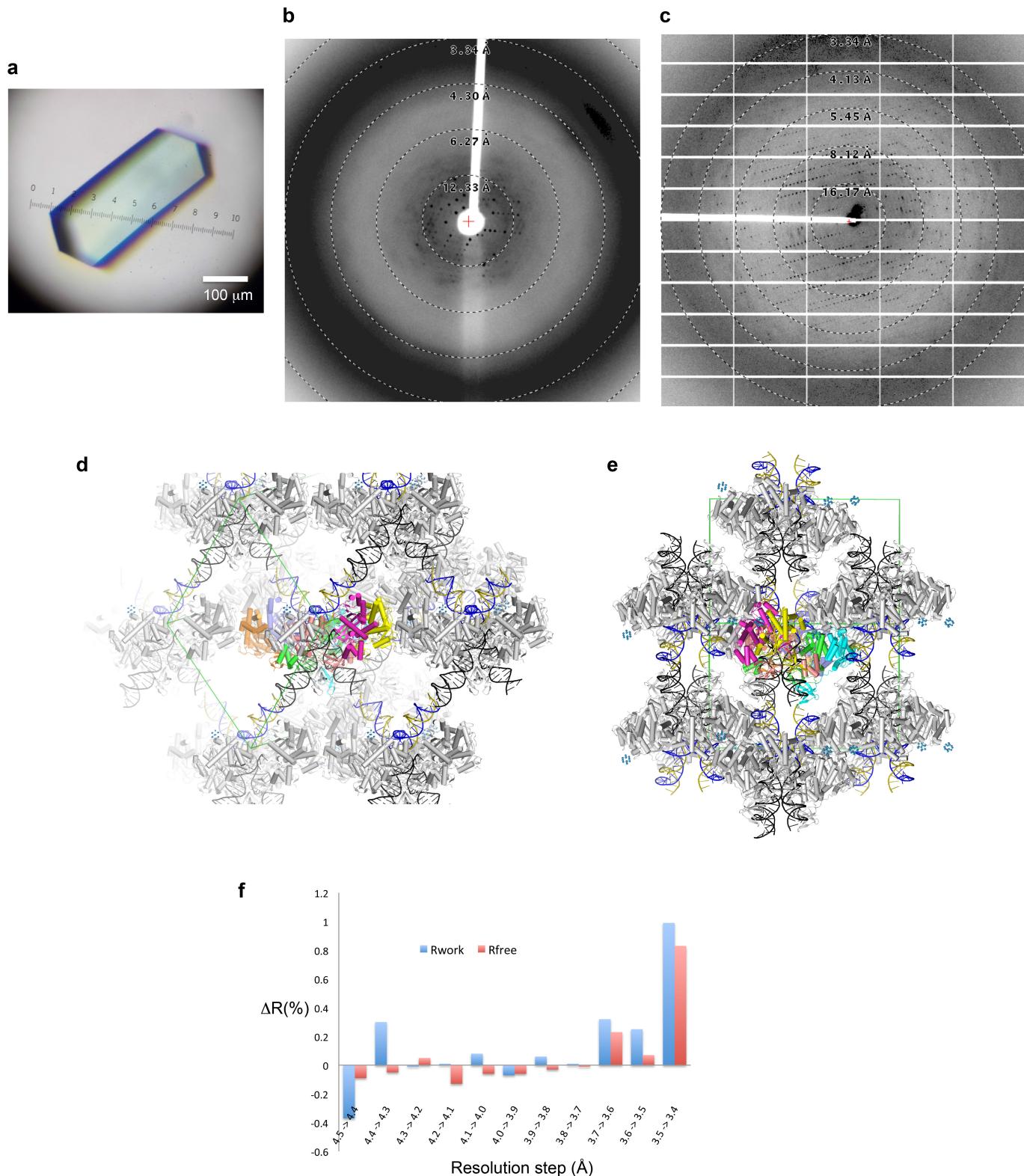


Extended Data Figure 1 | Amino-acid sequence alignment of RSV, HIV-1, and PFV INs. The secondary structure elements for RSV IN are colour coded on the basis of the three IN domains similarly to Fig. 1a. The residue numbering at the top is for RSV IN. For each IN, the black dots mark every ten residues. This figure was made using ESPript⁴¹.

a**b**

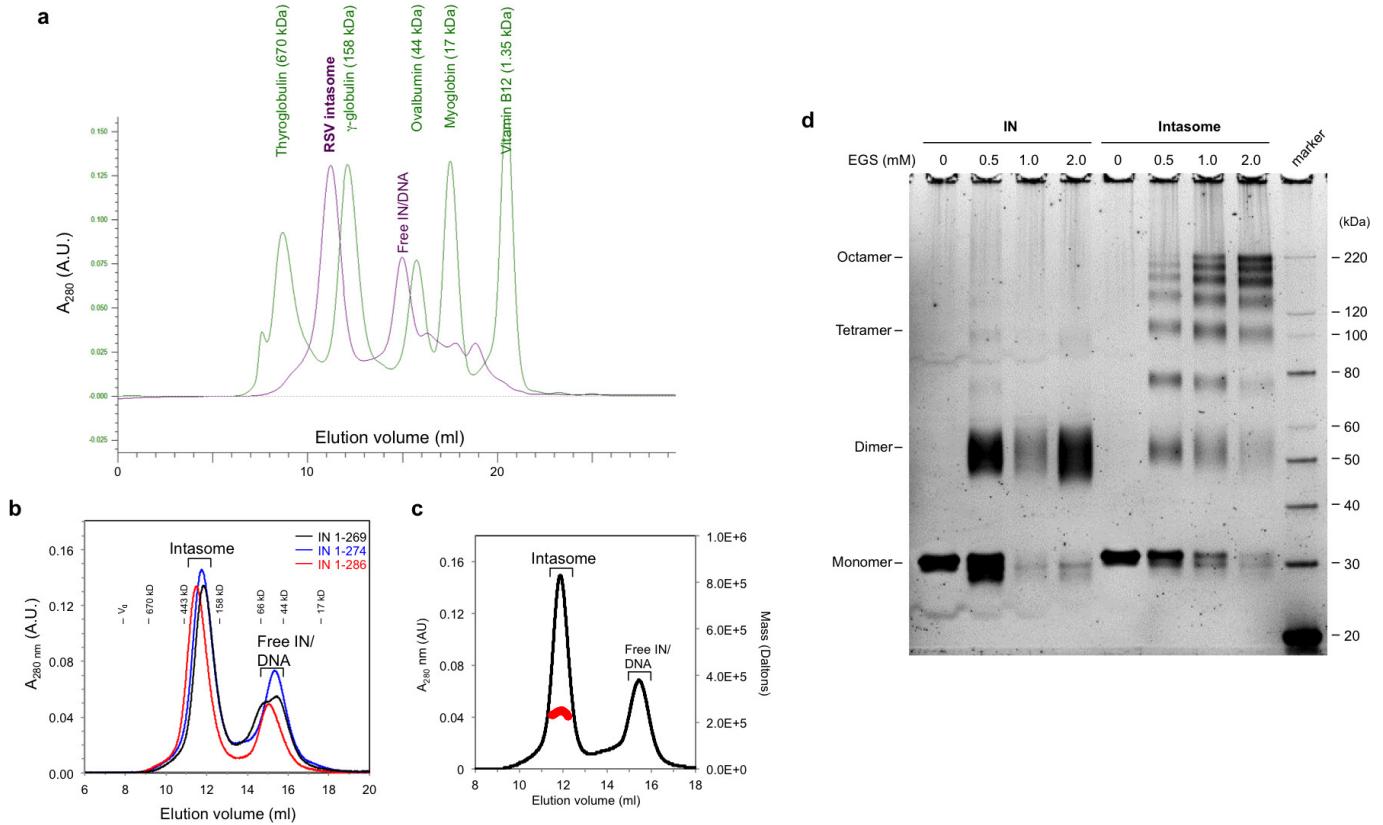
Extended Data Figure 2 | DNA substrate used for assembling the RSV intasome. **a**, The half-site (gapped duplex) substrate prepared by annealing three oligonucleotides dimerizes via the self-complementary six-base spacer sequence (underlined) to form a branched structure

mimicking the product of the concerted integration reaction (Fig. 1b). **b**, DNA structure in the RSV intasome. Viral DNA nucleotides are numbered, and some of the structural elements of RSV IN involved in the viral DNA interactions are shown.



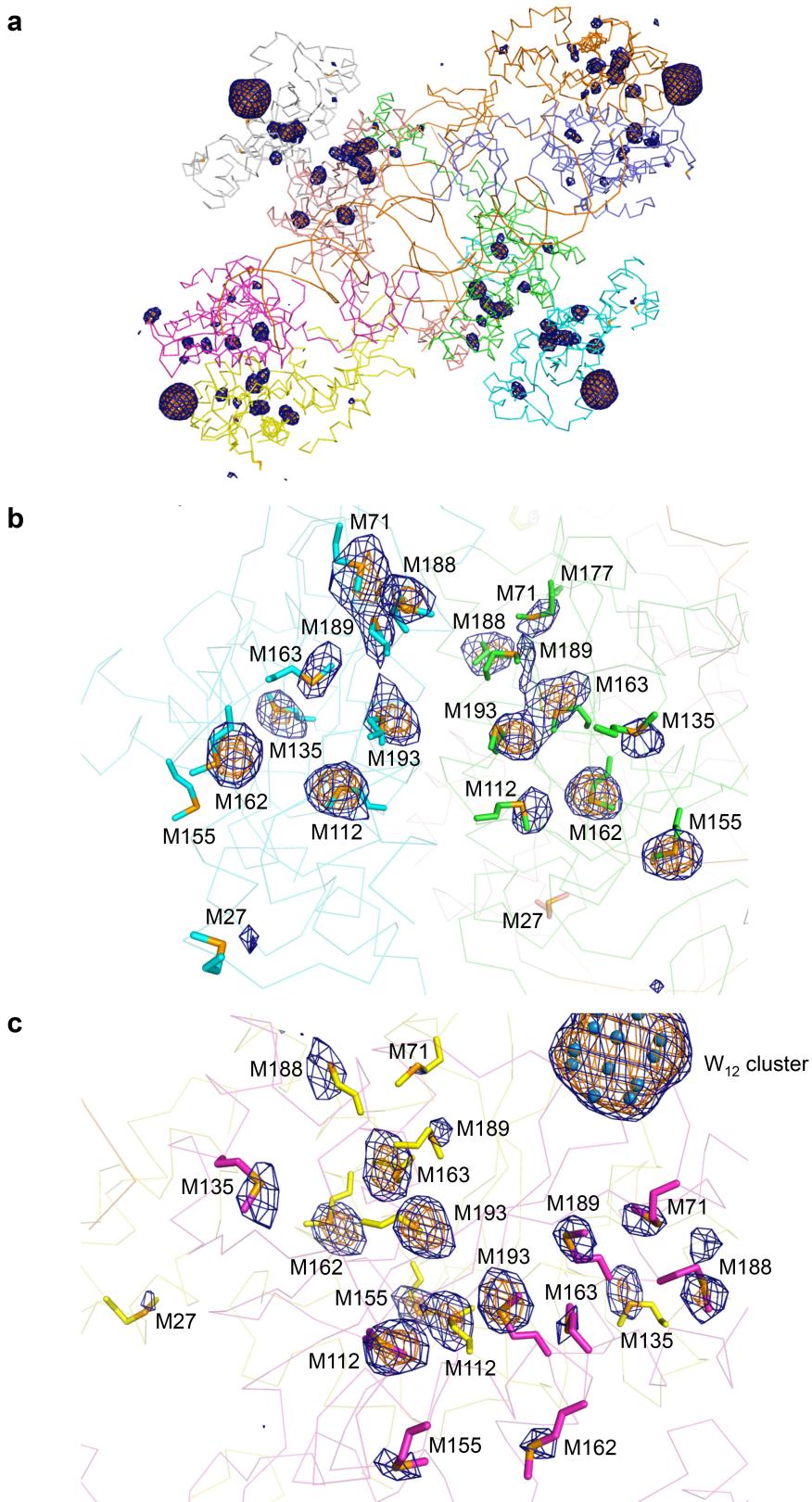
Extended Data Figure 3 | RSV intasome crystal. **a**, A crystal of the RSV intasome. **b**, X-ray diffraction pattern from a crystal not treated with metatungstate. **c**, X-ray diffraction pattern after the metatungstate-soaking (see Methods for details). **d**, **e**, Lattice contacts within the RSV intasome crystal. DNA strands are coloured as in Fig. 1b, while IN subunits from only one intasome are coloured. The unit cell is shown in green. The small blue spheres represent tungsten atoms. The view in **e** is perpendicular

to the twofold screw (*b*) axis of the monoclinic lattice, which lies horizontally. **f**, Paired-refinement analysis⁴⁰ to assess the resolution limit of the RSV intasome diffraction data. For each pairwise comparison, model refinements were run at two different resolution limits and the *R*-factors calculated for a common (lower) resolution cutoff were compared. Inclusion of data beyond 3.7 Å in the refinement compromised model quality.



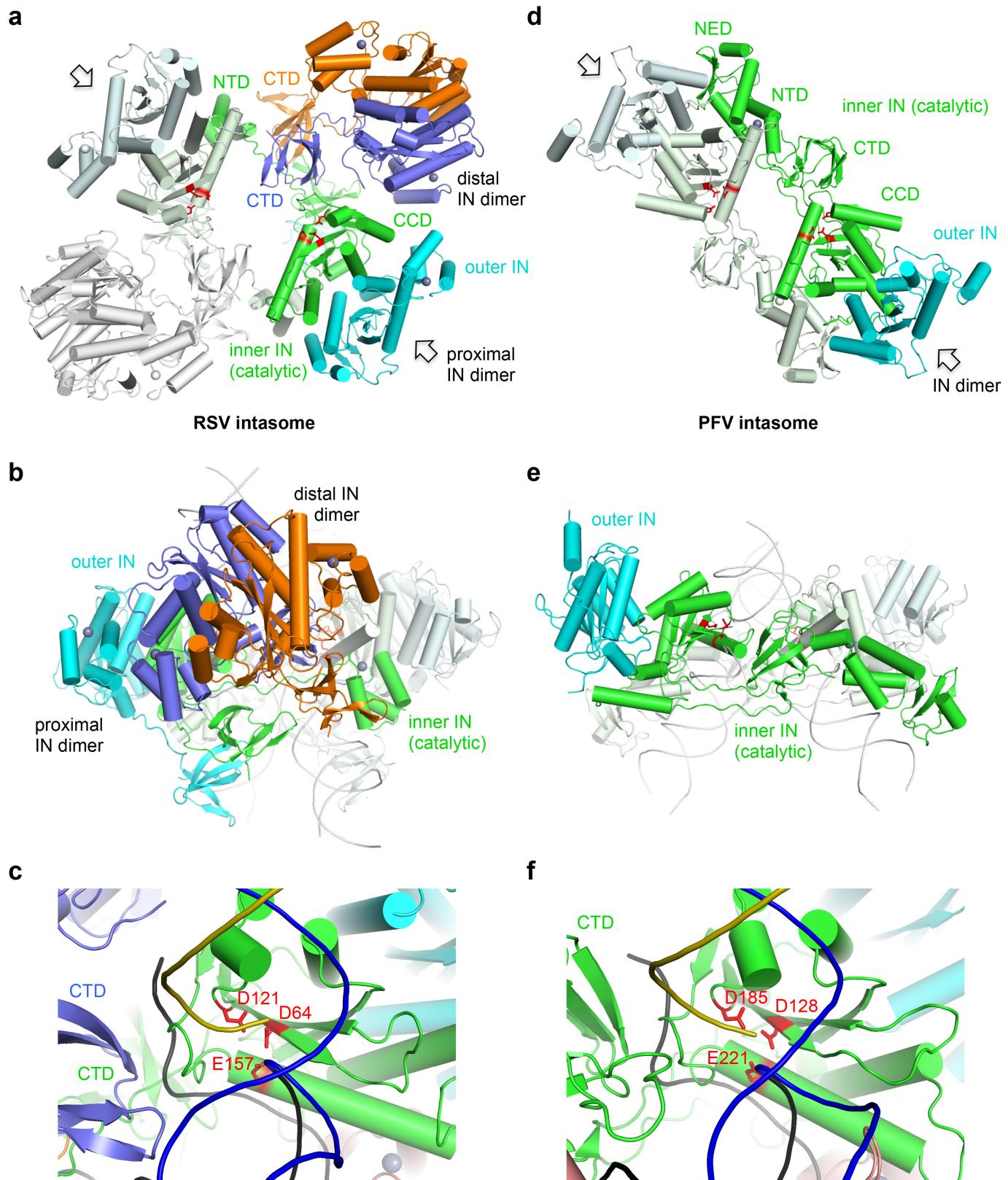
Extended Data Figure 4 | Biochemical characterization of RSV intasome. **a**, Representative SEC profile for RSV intasome, overlaid with that for a mixture of molecular mass markers. The buffer condition was as mentioned in the methods. **b**, SEC profiles for RSV intasomes formed with IN of varying C termini. **c**, SEC-MALS analysis of RSV intasome. The intasome formed with RSV IN (1–269 amino acids) was separated by SEC in a modified condition containing 20 mM HEPES pH 7.5, 1.0 M NaCl, 5% glycerol, and 1.0 mM TCEP. The absolute molecular mass was determined by light scattering using in-line detectors described previously⁴². The mass profile for the intasome is shown in red across the peak. The molecular mass of RSV intasome was 240 ± 10 kDa ($n = 4$). A similar SEC-MALS analysis of the intasome formed with wild-type full-length RSV IN (1–286 amino acids) yielded a molecular mass of 268 kDa ($n = 2$, data not shown).

The calculated mass of an intasome containing eight RSV IN (1–269 or 270) molecules is ~ 288 kDa. **d**, Chemical cross-linking analysis of RSV intasome. The RSV intasome and free IN (1–269 amino acids) were purified by SEC in the running buffer: 20 mM HEPES (pH 7.5), 1.0 M NaCl, 5% glycerol, and 1.0 mM TCEP. The peak fractions of the intasome and IN were cross-linked with the indicated amount of ethylene glycol bis-succinimidylsuccinate (EGS) as described previously⁴² and analysed by SDS-PAGE. Most cross-linked species within the intasome were larger than a tetramer. The highest oligomeric species observed is consistent with an octamer migrating at ~ 220 kDa. The molecular mass markers are in the far right lane. A NuPAGE 4–12% gradient gel with a MES-based SDS-PAGE running buffer was used.



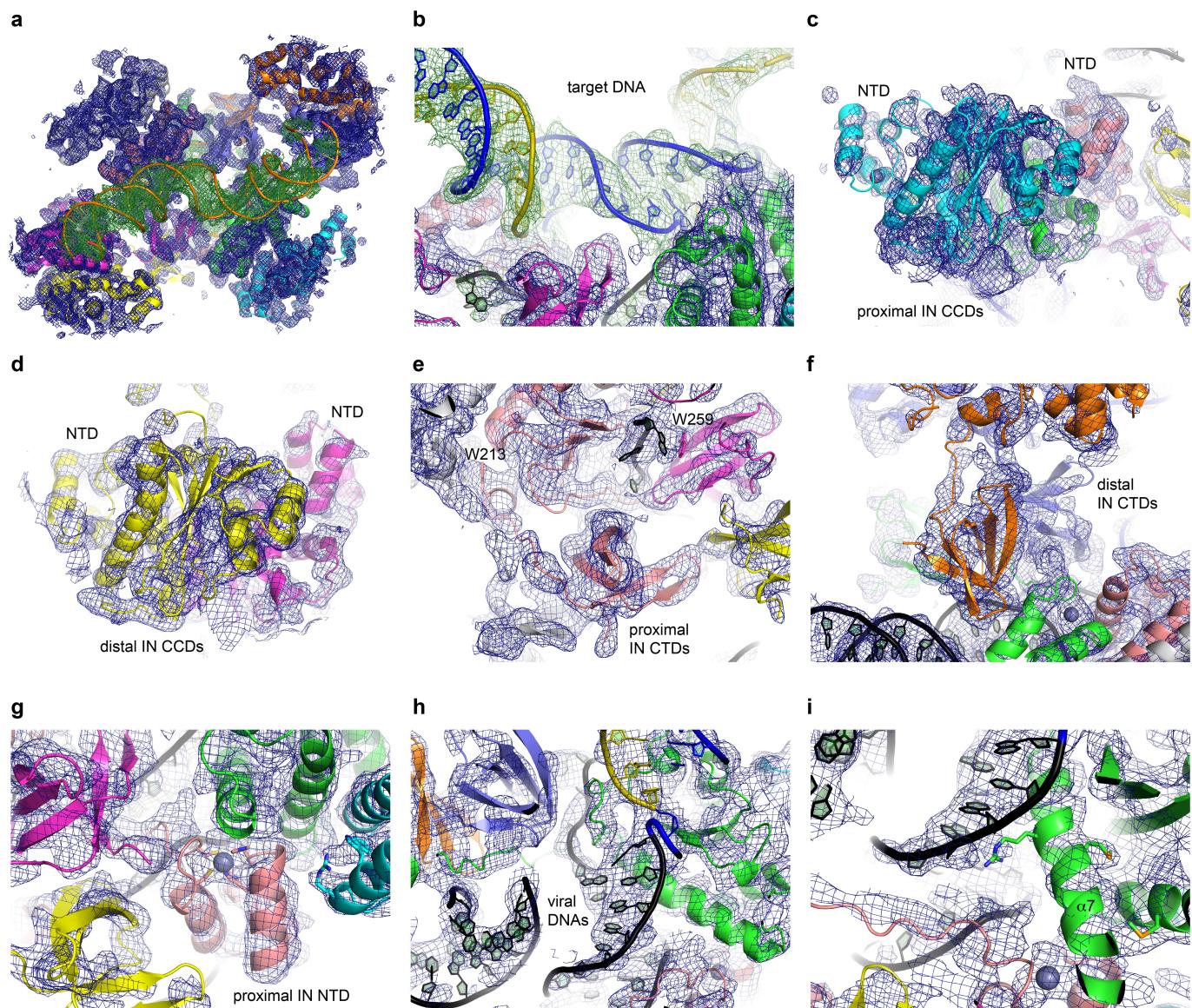
Extended Data Figure 5 | Selenium anomalous difference Fourier peaks confirming the model. Anomalous difference Fourier maps calculated using the data collected on selenomethionine-labelled RSV intasome at the Se K-edge wavelength, contoured at 3.5σ (blue mesh) or 5.0σ

(orange mesh). Methionine side chains are shown in sticks. **a**, A view covering the octameric RSV intasome. **b, c**, Close-up view of a proximal (**b**) and distal (**c**) IN dimer, respectively.

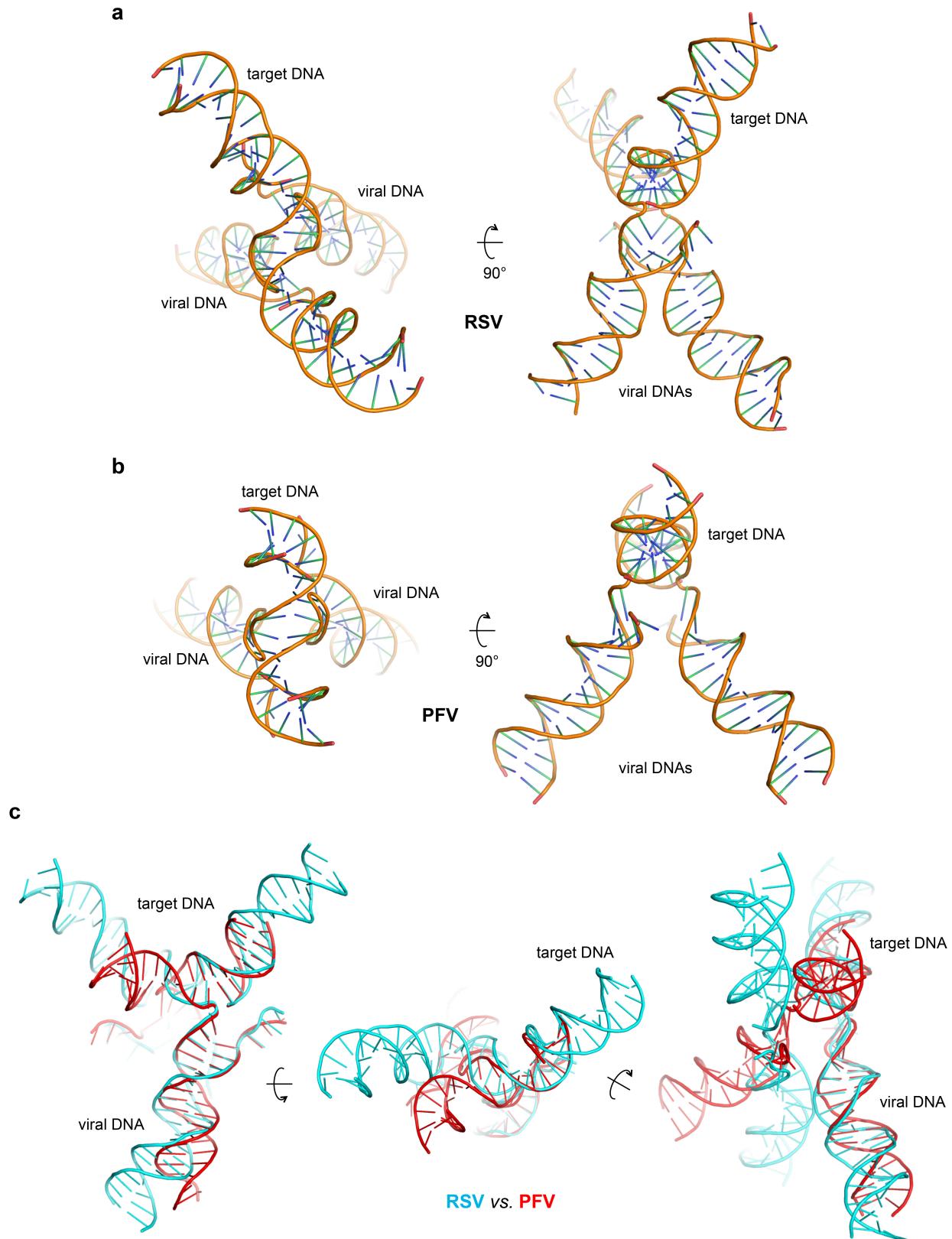


Extended Data Figure 6 | Comparison between RSV and PFV intasomes. **a, b**, Protein arrangement in the octameric RSV intasome. The inner and outer subunit of one proximal IN dimer is coloured in green and cyan, respectively, with the catalytic triad (DDE) of the inner subunit shown in red. The other proximal IN dimer is coloured similarly but in more pale colours. Arrows indicate the two proximal IN dimers. A distal IN dimer is coloured in slate and orange. DNA is omitted in **a, b**. **c**, Same view as Fig. 3c. **c**, Close-up view around the active site of the inner IN

subunit in the RSV intasome. The DNA strands are coloured as in Fig. 1, and the catalytic triad residues (DDE) are shown in red sticks. **d, e**, Protein arrangement in the tetrameric PFV intasome^{2,3}. The colour scheme follows that used for the proximal IN dimers of RSV IN in **a** and **b**. Arrows indicate the two IN dimers. DNA is omitted in **d, e**. **e**, Same view as Fig. 3f. **f**, Close-up view around the active site of the inner IN subunit in the PFV intasome (PDB accession number 3OS0 (ref. 3)). The colour scheme follows that in **c**.

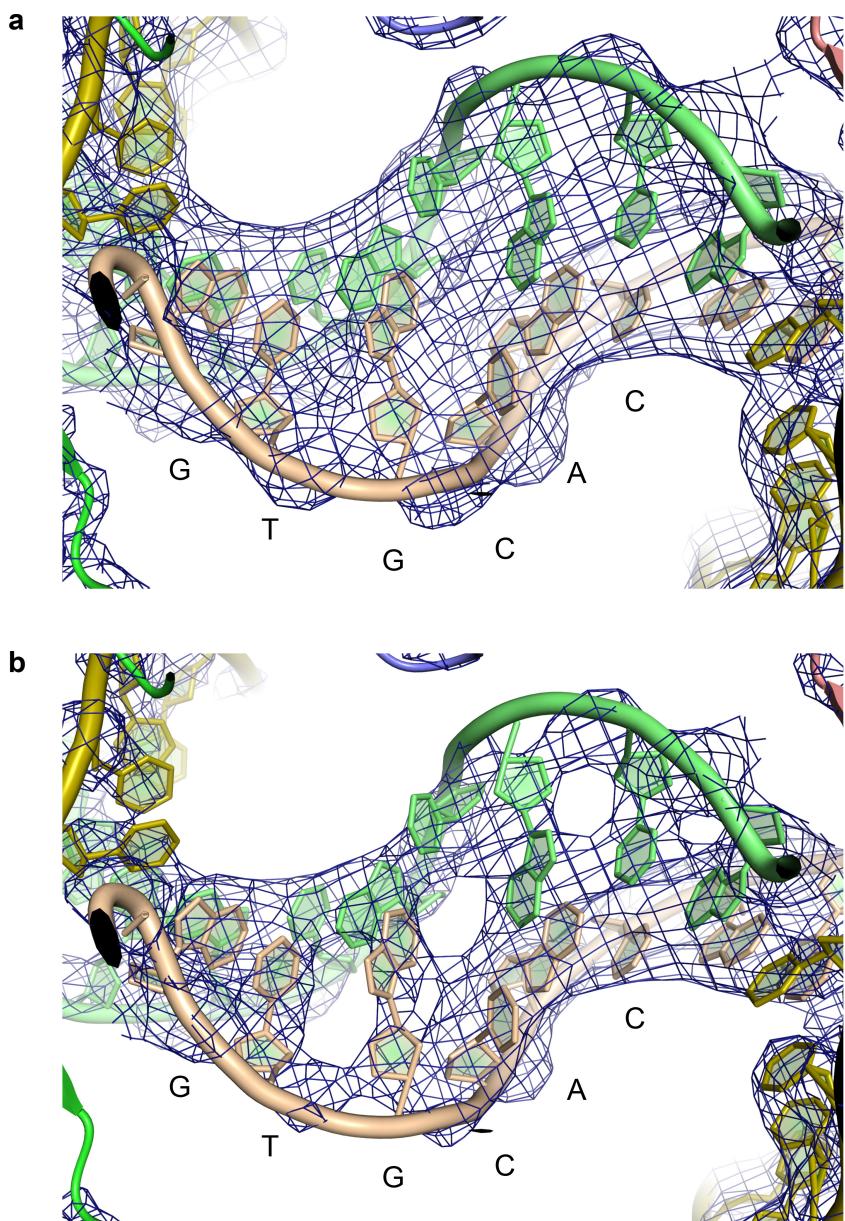


Extended Data Figure 7 | Composite omit maps. Simulated annealing composite omit $2mF_o - DF_c$ density contoured at 1.0σ , shown for area within 3.5 Å from any protein or DNA atom in the final model. Various parts of the RSV intasome are shown in panels a-i. In a and b, electron densities around protein and DNA are coloured differently (blue and green, respectively).



Extended Data Figure 8 | DNA conformations in the RSV and PFV intasomes. **a, b**, DNA structure in the RSV (**a**) or PFV (**b**) intasome, alternatively referred to as the strand-transfer complex (STC). The PFV intasome model is PDB accession number 3OS0 (ref. 3). **c**, A comparison of DNA structures between the RSV and PFV intasomes (STCs). The integration product DNAs (RSV in cyan, PFV in red) superimposed at

a viral DNA terminus are shown in three different view angles. Note the significant deviation in overall trajectory of the target DNA, and difference in the orientation of the second viral DNA molecule. The region spanning the two integration sites on opposing target DNA strands is 6 bp for RSV and 4 bp for PFV.



Extended Data Figure 9 | Electron density for the central 6 bp of the target DNA. The sigma-A weighted $2mF_o - DF_c$ map contoured at 1.5σ (a) or 2.5σ (b), overlaid with the final model for the central 6 bp region between the two integration sites.

Extended Data Table 1 | Data collection and refinement statistics

	RSV intasome
Data collection	
Space group	P2 ₁
Cell dimensions	
a, b, c (Å)	124.9, 157.8, 126.6
α, β, γ (°)	90.0, 110.9, 90.0
Resolution (Å)	49.3 - 3.80 (3.90 - 3.80)
R_{merge} (%)	14.7 (247.6)
$I/\sigma I$	6.0 (0.6)
CC _{1/2}	0.994 (0.287)
Completeness (%)	99.2 (99.3)
Redundancy	5.5 (5.5)
Refinement	
Resolution (Å)	49.3 - 3.80
No. reflections	44627
Reflections for R_{free}	2183
$R_{\text{work}}/R_{\text{free}}$	25.4/29.4
No. atoms	
Protein	19309
Ligand/ion	44
Water	0
B-factors	
Protein	206.9
Ligand/ion	205.0
Water	N/A
R.m.s deviations	
Bond lengths (Å)	0.004
Bond angles (°)	0.72

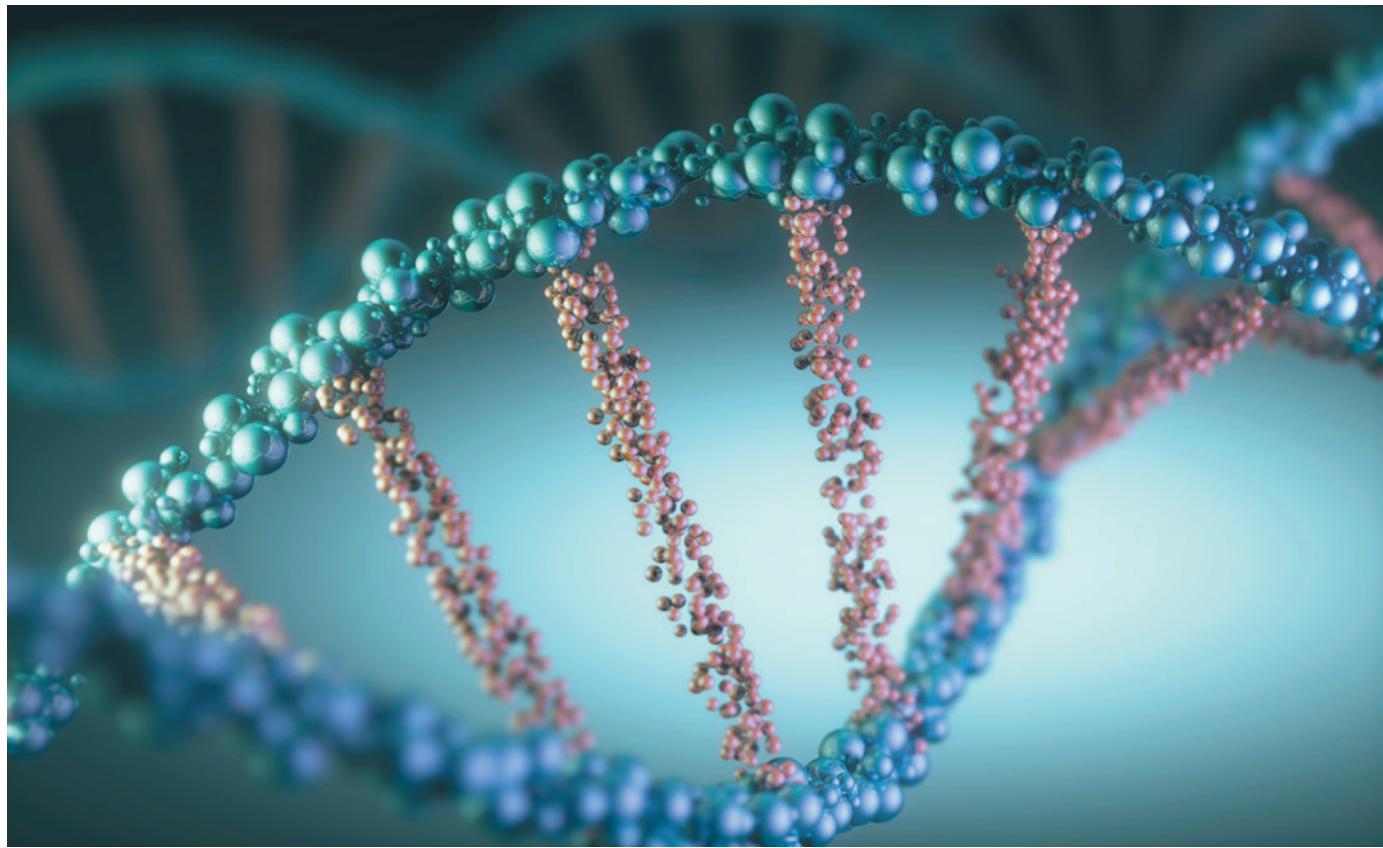
Statistics for the highest resolution shell are shown in parentheses.

TECHNOLOGY FEATURE

DNA TAGS HELP THE HUNT FOR DRUGS

Drug discovery is a daunting process that requires chemists to sift through millions of chemicals to find a single hit. DNA technology can dramatically speed up the search.

KTSDESIGN/SPL



BY ASHER MULLARD

Nestled in a plastic box, in an ordinary laboratory freezer on the second floor of a concrete building in Waltham, Massachusetts, is a clear test tube that contains a concoction of astronomical proportions. The library frozen within, a collection of chemical compounds owned by London-based pharmaceutical company GlaxoSmithKline (GSK), contains as many as 1 trillion unique DNA-tagged molecules — ten times the number of stars in the Milky Way.

This and other such libraries are helping pharma companies and biotechnology firms to quickly identify candidate drugs that can latch onto the proteins involved in disease,

especially those proteins that have proved difficult to target. They enable screening to be performed much more rapidly and cheaply than with conventional methods. And academic scientists can also use them to probe basic biology questions and investigate enzymes, receptors and cellular pathways.

Drug discovery often starts with researchers assembling large libraries of chemicals and then testing them against a target protein. Compounds are added individually to wells that contain the target to see whether they affect its activity. This approach, known as high-throughput screening (HTS), is automated using robotic equipment to test millions of chemicals, but it's still laborious, expensive and not always successful.

Over the past few years, medicinal chemists have been increasing the odds of finding potentially useful compounds by labelling chemical compounds with bits of barcode-like DNA. These DNA-encoded libraries — which can dwarf conventional small-molecule libraries — offer all sorts of advantages to drug discovery. For a start, rather than testing each compound individually, researchers can put all of the DNA-tagged small molecules into a single mixture and then introduce the target protein. Any compounds that bind with the target can be identified easily thanks to their DNA barcodes.

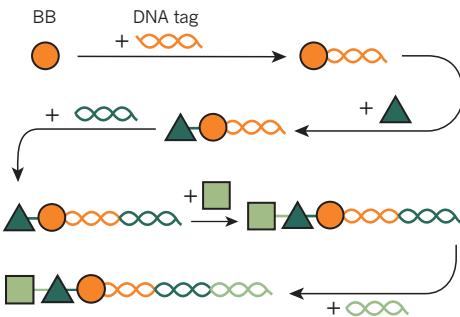
DNA-encoded libraries were first proposed in 1992, in a thought experiment¹ by molecular biologist Sydney Brenner and chemist Richard

BUILDING BARCODES

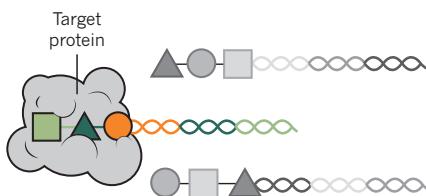
DNA-encoded libraries consist of billions of molecules tagged with identifying DNA 'barcodes'. The biggest libraries are built using DNA recording, but DNA templating allows for finer control.

DNA recording

1 ► A building block (BB), such as an amino acid, is tagged with a DNA sequence. Then a second BB is added, and the DNA tag is lengthened with a sequence that corresponds to the second BB. Compounds of up to four BBs can be assembled. At each step, the BBs and tags are added to mixtures of thousands of different tagged BBs to quickly grow libraries to vast sizes.



2 ► A target protein is added to the mixture of billions of unique DNA-tagged molecules, some of which bind to the target. The DNA barcodes of the bound molecules are used to identify them.

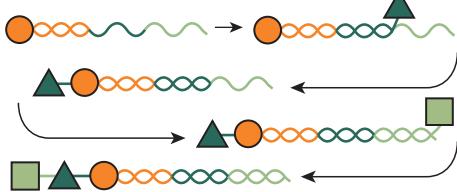


DNA templating

1 ► Each molecule is designed as a single-stranded DNA template. BBs are tagged with DNA 'anti-barcodes' that are complementary to regions on the planned molecule's template.



2 ► A DNA-tagged BB binds to its corresponding section on the template. A second DNA-tagged BB is added and binds to its corresponding template position, and the two BBs join in a chemical reaction. More BBs are added to complete the template.



3 ► A final chemical reaction can convert a string of building blocks into a ring, producing barcoded macrocycles.



Lerner, who were then at the Scripps Research Institute in La Jolla, California. They have been gaining momentum ever since. In 2007, GSK acquired one of the firms that pioneered these libraries, Praecis Pharmaceuticals in Waltham, for US\$55 million. The drug firms Novartis and Roche, both in Basel, Switzerland, have started their own in-house DNA-encoded-library programmes. A burgeoning group of biotech companies — including X-Chem in Waltham; Viperen in Copenhagen; Ensemble Therapeutics in Cambridge, Massachusetts; and Philochem in Zurich, Switzerland — has meanwhile built up a who's who list of industry and academic partners that are eager to use the technology.

"People understand now that this is not a fad," says Robert Goodnow, executive director of the Chemistry Innovation Center at AstraZeneca in Boston, Massachusetts, which collaborates with X-Chem. "It's for real."

DNA-encoded libraries will not replace HTS: companies have already invested heavily in HTS screening, and there are some compounds that cannot be synthesized using DNA-encoding technologies. Rather, they offer a complementary way to quickly, efficiently and cheaply find chemical structures that bind to new or historically challenging targets, such as ubiquitin ligases, which flag proteins for disposal and could be targeted in cancer therapy.

BIG IS BEAUTIFUL

GSK currently has the world's biggest DNA-encoded library: it is an impressive 500,000 times larger than the company's 2-million-compound HTS library.

There are several ways to build DNA-encoded libraries: the biggest ones, like GSK's, are made using an approach called 'DNA recording' (see 'Building barcodes'). Chemical building blocks, such as amino acids, amines and carboxylic acids, are synthesized and then tagged with a unique DNA barcode through a chemical reaction. A second building block is added to the mix to make a new small molecule, and the DNA barcode is then lengthened. By joining up to four blocks, chemists can create drug-like molecules. And because they have thousands of building blocks to play with, the number of potential combinations is enormous.

Compared with conventional HTS libraries, for which chemists have to test each compound individually, DNA-encoded libraries are easier to maintain and use. A DNA-encoded library can be stored in a single test tube, whereas an HTS library requires robot-filled facilities that are large enough to store each compound individually.

But the true beauty of DNA-encoded libraries, says Chris Arico-Muendel, a manager at GSK in Waltham, is the sheer number of chemical structures it is possible to synthesize. The company's drug-discovery team now

uses the DNA-encoded library as frequently, if not more frequently, than the HTS library for new and difficult protein targets. The most advanced compound to emerge from the company's DNA-encoded library so far is GSK2256294, which blocks epoxide hydrolase², an enzyme that is involved in breaking down lipids. This drug candidate came out of GSK's collaboration with Praecis and has completed first-in-human safety studies that may support further evaluation of its use in diabetes, wound healing or as a therapy for chronic obstructive pulmonary disease³. "We are pleased with how things are going with DNA-encoded libraries within GSK," says Arico-Muendel.

And as more chemical building blocks are created, along with extra ways to attach them to one another, the breadth of these libraries will continue to expand.

In the near future, DNA-encoded libraries will not only become bigger and broader but might also provide hits that can quickly be moved into the clinic, says X-Chem chief executive Richard Wagner. With conventional screening, medicinal chemists sometimes have to spend many years tweaking compounds to make them specific, potent and safe enough to enter the clinic. "This is just a game of odds," says Wagner. By contrast, the large size of DNA-encoded libraries means that, by chance, some of the compounds they include will be more clinic-ready than others. Although the compounds will still require optimization, "we can get things that are pretty close", he says.

X-Chem, which holds 120 billion compounds in its DNA-encoded library, is already starting to see this in practice. It took just one year to move its most advanced candidate — an autotaxin inhibitor that blocks the conversion of one phospholipid into another — from a screening hit to a clinical candidate. A spin-off company of X-Chem, X-Rx in Wilmington, North Carolina, now plans to start clinical trials of the compound for fibrosis in 2017. Interest in X-Chem's library is spreading across the industry: in the past five years, the company has forged collaborations and licensing agreements with several major pharmaceutical firms — including Roche, AstraZeneca, Bayer, Johnson & Johnson, Pfizer and Sanofi — as well as with a host of biotech and academic partners.

MADE TO ORDER

Other biotech firms have added an interesting twist. They use the DNA tag not only to identify a compound but also as a template to make it. David Liu, a chemist at Harvard University in Cambridge, and his students developed this 'DNA-templated' approach and used it to build a library of circular molecules called macrocycles⁴. These larger, more-stable, ring-shaped molecules interact with the target at multiple sites, boosting the specificity of the binding reaction. (GSK and X-Chem also have

extensive macrocycle collections in their DNA-encoded libraries.)

Liu first creates single-stranded DNA templates that act as guides — these consist of several regions that are complementary to the DNA tags on his chemical building blocks. He then sequentially adds the DNA-tagged building blocks into a reaction vessel, relying on DNA base pairing to physically bring the tagged building blocks close enough together to bind to one another. A final reaction then converts the strings of building blocks into rings, producing macrocycles that are each tethered to a unique DNA barcode.

Constructing a DNA-templated library involves a hefty workload because researchers must design a template for each molecule as well as tagging thousands of building blocks with DNA. As a result, DNA-templated libraries are smaller than DNA-recorded ones, but they still eclipse HTS libraries in terms of size — and they have other advantages, too. Because scientists know at the outset what compounds they are producing, they can purify the DNA-templated libraries to remove compounds that are tagged inaccurately. This step translates into a high degree of confidence in the hits. By contrast, colossal DNA-recorded libraries may still contain wrongly tagged compounds, and thus might yield hits that will send researchers on wild goose chases.

Liu's 14,000-strong library has already led to a few triumphs. In 2014, his team reported that it had solved a problem that researchers had been struggling with for decades when it found⁵ a specific and stable small molecule that can block insulin-degrading enzyme (IDE), which has been linked with type 2 diabetes. He and others have started to unravel the role of IDE in both health and disease, which has led to the identification of other IDE inhibitors. Discussions are under way to develop these into drugs.

Liu has also screened more than 100 other targets, many brought to him by academic collaborators who need small-molecule inhibitors of their pet proteins. "I never would have thought seven years later that this first-generation library would still be providing us with interesting biological discoveries," he says. "But it has proved to be very fruitful. We have had more hits against targets from our first library than we can follow up on."

He is nevertheless putting the finishing touches to a second-generation, 256,000-macrocycle library that could open up even more biology. Ensemble Therapeutics, which Liu founded in 2004, now has more than 10 million macrocycles in its library. The company is focusing on targets among the immune checkpoint proteins, which modulate

"We have had more hits against targets from our first library than we can follow up on."



Sydney Brenner, who conceptualized DNA-encoded libraries with Richard Lerner in 1992.

the immune system, and the ubiquitin ligases. It has also granted a license to Novartis to develop one of its finds, a molecule that targets the inflammation-linked protein interleukin-17.

SWEET SCREENS

Once the library is built, the fun of identifying which molecules stick to a target begins. Most researchers rely on 'affinity-based screening' to find those compounds. For this, they engineer the protein target to include a purification tag. They then pass the mixture containing the library and target through a purification column, using the purification tag to pull out the bound pairs. The last step is to read the DNA identifiers linked to the small molecules using a DNA sequencer.

This method can yield results even with minute amounts of a target protein. In one project, remembers Arico-Muendel, academic collaborators wanted to screen an unstable protein that they could produce only in tiny quantities. "They flew it here on dry ice overnight, and we immediately did the entire screen on it," he says. "And that actually gave some really good hits." Such experiments are impossible with HTS, because the target protein must be stable and abundant enough to be added into millions of wells before the experiment can begin.

But affinity-based screening has its shortfalls. The chunky DNA tag can sometimes impede interaction with the target, and some potential candidates may be lost. But because the DNA-encoded libraries are so big, screeners are not typically too concerned with these losses. More problematically, small molecules and their tags can bind to the purification column and generate false-positive hits. The purification tag can also interfere with the structure of the target protein, introducing confusion in the data.

Several groups have developed solutions for this problem. Vipergen, a biotech firm that

has a DNA-templated library with 50 million molecules, has put its hope in a 'binder trap enrichment' strategy.

Imagine, says Nils Hansen, the company's chief executive, that you could freeze your protein-library mixture and cut it into super-small ice cubes. If the ice cubes are small enough, each will be able to contain only a single target protein. At this size, small molecules that bind to the target will be consistently overrepresented in ice cubes that contain targets, even without a purification strategy. Vipergen has achieved the same effect by putting its screens into water-and-oil emulsions, in which minuscule water droplets stand in for ice cubes⁶. "It's pretty cool," says Hansen.

Currently, screens of DNA-encoded libraries work best with free-floating, soluble proteins. But many appealing drug targets are embedded in the cell surface, making them impossible to probe with traditional affinity-based screening. For example, an estimated 40% of approved drugs target membrane-bound G-protein-coupled receptors, which sense molecules outside the cell. The technology for screening membrane-bound proteins is evolving, says Goodnow, "but is still kind of a challenge".

One way forward is to mix a DNA-encoded library with intact cells that overexpress a membrane-bound target. The small molecules can then bind to the targets on the surface of the cell. After the researchers wash away the unbound library, they can identify the bound small-molecule hits by heating up the cells and reading the eluted DNA tags. GSK has used this approach to identify potent inhibitors of a receptor that has been implicated in schizophrenia and disorders of the central nervous system⁷.

X-Chem, too, is starting to see success with screens of membrane-bound proteins. "Historically, the majority of our programmes were on soluble proteins. But there is shift given the recent data we've been able to generate with really difficult membrane-bound proteins," says Wagner.

With DNA-encoded libraries continuing to expand, and new screening approaches opening up uncharted biological space, he adds, "DNA-encoded libraries are set to become one of the pillars of discovery in the pharmaceutical industry." ■

Asher Mullard is a freelance science journalist in Ottawa, Canada.

1. Brenner, S. & Lerner, R. A. *Proc. Natl Acad. Sci. USA* **89**, 5381–5383 (1992).
2. Clark, M. A. *et al.* *Nature Chem. Biol.* **5**, 647–654 (2009).
3. Laza, A. L. *et al.* *Br. J. Clin. Pharmacol.* <http://dx.doi.org/10.1111/bcp.12855> (2016).
4. Tse, B. N., Snyder, T. M., Shen, Y. & Liu, D. R. *J. Am. Chem. Soc.* **130**, 15611–15626 (2008).
5. Maianti, J. P. *et al.* *Nature* **511**, 94–98 (2014).
6. Blakskjaer, P., Heitner, T. & Hansen, N. J. *V. Curr. Opin. Chem. Biol.* **26**, 62–71 (2015).
7. Wu, Z. *et al.* *ACS Comb. Sci.* **17**, 722–731 (2015).

CHRYsalis

Flight into danger.

BY ANDREA KRIZ

I remember the broadcasts. *BREAKING NEWS: EXPLOSIVE END FOR PROJECT CHRYsalis*. An air-force pilot lowering herself into the cockpit with a thumbs-up and a smile. Nursing black eyes and bruises in the Earth Archives, I glued myself to the screen a thousand times as she unravelled, as glass melted into a membrane that engulfed her in blinding light. And I recall the curses behind my father and me in the Afernsi refugee ship's mush line. "We never should've toyed with it. It was never meant for us. Charlie Bickoid — your curiosity killed us all."

My father always spoke of it like some kind of fairy tale with a moral. Project Chrysalis, humankind's attempt to catch up in the galactic arms race. To reverse-engineer an alien 'ship' that fell from the heavens. The biochemists purified a protein-like macromolecule from its circuits. Dismissed it as contamination. An enzyme, soon discovered. A life form. When the Sopholid pupa exploded in the US Air Force Test Center, it sent out some kind of signal. *THIS JUST IN*. Two wings unfurling out of the inferno.

I look up at the dogfight and wonder if he's up there. If that's his plane supernovaing in the violet skies, torn in two by the swarm. My father. He's decades too old to be crunched up in a turret ball. He's a neuroscientist. It should be me. But I've no sense of duty, I tried to run from the draft and they dragged me to this, humanity's last refuge of a colony, kicking and screaming.

Like the Sopholid that Buffalo zapped, writhing beside me. A Mantis. Blood still streams off its scythes. The remains of its core drip into the bone-white rock at my feet. I wonder if it used to be human. These gleaming orbs — their Queen's heart throbs overhead — reason enough, apparently, to hound us to the far edge of the Galaxy, to plague our colonies until each and every one of us drools lobotomized and domesticized on one of their farms. "Condensed organic material," was all Commander Xexe would tell us in our briefings. The Sopholids have strengthened a hundredfold since developing human-powered cores, the Afernsi added grimly. It was good of the Afernsi to rescue a couple of hundred thousand of us from Earth if that's the case, I suppose. As the last troopship fades into the atmosphere,



I wonder how many they've deigned to save this time.

"Now."

Buffalo's shaking. His eyes dart between the Locusts gnawing on the rest of our squad and the Queen above, coiled around the hover-carrier of evacuees we were supposed to protect.

"I'll fly up there and tear her a new one," I whisper, crouching by the Mantis's head. "And if the Queen shuts down, the rest of the swarm should retreat, right?"

"I'll shoot you if you touch that thing," Buffalo mumbles.

He wants me to remember what happened the last time a human wired herself into a Sopholid's brain. Biomechanical equations that've been worked out in excruciating detail since. The raw energy of metres of DNA uncoiling from each and every one of the trillions of cells in the human body simultaneously. Amplified by the circuits of the Sopholid in a way we still don't fully understand — such was the topic of my PhD research that I hope to, no, that I *will* return to after the war — it was akin to a nuclear bomb.

"Are you saying you want me to just stand here and die?"

► **NATURE.COM**

Follow **Futures**:

Twitter: **@NatureFutures**

Facebook: **go.nature.com/mtoodm**

He says nothing. But I know. His Afernsi ray gun only fizzes when he presses

the fire button. All we've got left are our good ol' Earth rifles, whose bullets bounce off Sopholid exoskeletons like staples.

A sound like a beetle popping underfoot rings out. Buffalo crumples, clutching his side.

"Turbo Eater," he finally manages. "You really believe humans can fly?"

"I don't know," I admit. "We can pilot ships and things."

Not just to this miserable, terraformed rock that the nearest star hasn't risen on for days. But Satellite 10388. The Galactic University. My home. And the Afernsi refugee ship that brought me there. From Earth, but I don't remember any of it. I remember biting into a hydroponic tomato I stole when I was eight, the skin bursting, the juice dribbling down my lips. The constant ache of hunger. The stench. My dad and I lived in 20 square feet of space behind a radioactive waste container. And to conserve oxygen, they'd only let us open the hatch to roll out bodies twice a week.

"That's how we got here, right?"

I crawl forward. With that movement I affirm that I am, with all my heart, a believer in the Pilot Hypothesis. My father's, Charlie Bickoid's, fatal idea that someone built them. The Sopholids. That they can be controlled. I pry my fingers into the Mantis's skull — cockpit — curl up in the wetness, let the translucent plates close behind me. Brain mass glows to life around me. Spidery characters flash before my eyes. But I swipe all the warnings aside. Tendrils feel me out. I shudder as they burrow into my spine.

The film of the canopy was cracked and bloody, but suddenly I can see through it, somehow. I clench my fist and the scythe of the Mantis moves with me. And the Queen turns to look down at me at last. We thought their eyes were blank, compound, insectoid. But suddenly I can see all the colours of another rainbow in there. A rainbow I couldn't even have begun to comprehend before. They're calling me. She's calling me to join her.

"Turbo Eater." The wind whistles in Buffalo's mouth because I'd pulled the trigger first, I'd shot him hours and hours ago. "Fly." ■

Andrea Kriz is a PhD student in biological and biomedical sciences at Harvard University. She is the winner of the 2015 Ilona Karmel Prize for Writing Science Fiction.

ILLUSTRATION BY JACEY